Thermal adaptation of conformational dynamics in ribonuclease H

Kate Stafford

Submitted in partial fulfillment of the

requirements for the degree

of Doctor of Philosophy

in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2013

ABSTRACT

Thermal adaptation of conformational dynamics in ribonuclease H

Kate Stafford

Structural changes are critical to the ability of proteins, particularly enzymes, to carry out their biological function. However, flexibility also leaves proteins vulnerable to denaturation and degradation; thus a balance must be struck between the dynamics required for function and the rigidity required for maintaining a globular protein's characteristic folded structure. These relationships have been studied in detail through comparison of homologous proteins from organisms adapted to varying properties of the bulk environment. In particular, organisms adapted to temperature extremes offer fruitful platforms for the investigation of adaptive changes in protein stability as a function of environmental pressures. Thermostable proteins are widely reported to be more rigid than their homologs from mesophilic organisms, and those from psychrophiles more flexible; this suggests the possibility of evolutionary conservation of the balance between dynamics and stability. Thus specifically functional aspects of protein dynamics may be isolable through the comparative analysis of members of protein families from organisms adapted to different thermal environments.

The best experimental tool for characterizing internal conformational dynamics of proteins on a range of timescales and at site-specific resolution is nuclear magnetic resonance (NMR) spectroscopy, which has found widespread use in the study of protein flexibility and dynamics. However, it is often difficult to provide a detailed structural interpretation

of NMR observations. This gap can be bridged using molecular dynamics (MD) simulations, which can directly simulate motional processes that have been observed experimentally. The potential for deep synergy between these two complementary tools has been recognized since MD methods were first applied to biological macromolecules, and recent technological developments have reinforced the mutually beneficial relationship between the two techniques.

Ribonuclease HI (RNase H), an 18 kD globular protein that hydrolyzes the RNA strand of RNA:DNA hybrid substrates, has been extensively studied by NMR to characterize the differences in dynamics between homologs from the mesophilic organism *E. coli* and the thermophilic organism *T. thermophilus*. However, these dynamic differences are subtle and difficult to interpret structurally. The series of studies described in the present work was conceived in the pursuit of an improved understanding of the complex relationships between protein dynamics, activity, and thermostability in the RNase H protein family. The organizing principle of the work presented herein has been the close coupling between molecular dynamics simulations and NMR observations, permitting both validation of the MD trajectories by rigorous comparison to experiment and improved interpretation of the dynamics observed by NMR. Previous NMR observations of *E. coli* and *T. thermophilus* are integrated into an interpretive framework derived from simulations of the larger RNase H family.

First, comparative analysis of molecular dynamics simulations of a total of five homologous RNase H families from organisms of varying preferred growth temperature reveals systematic differences in the conformational dynamics of the handle region, a loop previously identified as contributing to substrate binding. Second, analysis of the effects of activating mutations on the dynamics of ttRNH identifies rotamer dynamics whose contributions to increased catalytic activity can be rationalized in the context of observed differences in sidechain orientation in the wild-type ecRNH and ttRNH simulations. Third, a combined MD-NMR study finds that the active site residues of ecRNH, and likely of the entire RNase H family, are rigid on the ps-ns timescale while undergoing substantial conformational exchange upon Mg^{2+} binding; this suggests that the active site is electrostatically preorganized for binding the first metal ion, which in turn induces dynamic reorganization at longer timescales. Finally, long-timescale simulations of the RNase H family, despite unexpected local unfolding for some family members, identify handle-loop and rotamer preferences for the *C. tepidum* RNase H (ctRNH) homolog that unexpectedly differ from those observed for ecRNH and ttRNH, and which can be experimentally tested by NMR spectroscopy of this recently characterized and less well-studied example of an RNase H homolog from a thermophilic organism.

Table of Contents

List of Figures

List of Tables

Acknowledgments

I would like to take this opportunity to thank the many people who have contributed to my intellectual development over the past five years for their time, expertise, and support. None of this work would have been possible in their absence.

First and foremost I would like to express my gratitude to Art Palmer, whose advice and mentorship have been crucial to my success as a scientist. With a finely tuned balance between *laissez-faire* laboratory management and uncompromising support at critical junctures, Art's guidance has been enormously beneficial, not only in terms of science but also as reminders to look outside the lab (or in my case, at something other than a terminal window), at least long enough to read enough of the New York Times to participate in our inevitable pre-lab-meeting conversations.

Secondly I thank the members of my permanent committee, Barry Honig and Wayne Hendrickson, who have provided helpful comments and support throughout the progress of this project, and the members of my defense committee, David Eliezer and David Shaw, who generously agreed to participate in grilling me about my work. Special thanks are owed to David Shaw, with whom I worked at D.E. Shaw Research prior to coming to Columbia, whose enthusiasm for molecular dynamics simulations proved infectious.

The members of the Palmer laboratory deserve thanks for their contributions to this project in particular and for their general ability to maintain a productive, friendly, and fun work environment. I hope I haven't missed anyone I overlapped with during my time in the group, but I offer many thanks to Jeff Chang, Jae-Hyun Cho, Bob Geis, Michelle Gill,

Cornelia Haas, Paul Harvilla, Ying Li, Paul O'Brien, Nichole O'Connell, Paul Robustelli, Nikola Trbovic, and Tim Zeiske. In particular I thank Nikola Trbovic for blazing the trail on applications of molecular dynamics simulations to the RNase H project; Jae-Hyun Cho and Paul Robustelli for their collaborative efforts on RNase H; Michelle Gill and the three Pauls—Harvilla, O'Brien, and Robustelli—for assistance in the wet lab (and I hope Paul Robustelli, my fellow computationalist, enjoys being acknowledged for such a thing); and Michelle Gill and Ying Li for their assistance with NMR experiments (and particularly their help in differentiating problems due to an inexperienced spectroscopist at the controls from problems due to something, somewhere, being broken). It's been a great five years working with everyone and I look forward to seeing the exciting science that will emerge from the Palmer lab in the future.

Several institutions also deserve acknowledgment for their support of my doctoral research. I was fortunate to have received a National Science Foundation Graduate Research Fellowship which supported me for the final three years of my Ph.D. The work presented here made extensive use of the resources of Columbia's Center for Computational Biology and Bioinformatics, and I thank their staff, particularly John Wofford, for tolerating my sometimes picky and unusual computing needs. Additionally, I thank the Pittsburgh Supercomputing Center and D.E. Shaw Research for the generous grant of computing time on the Anton platform.

To anyone who is reading this document, you too deserve acknowledgment for doing so. Especially if you are reading this section, as this writing marks the first time I've personally read the acknowledgments in the theses by Nikola Trbovic and Joel Butterwick, both of whom preceded me on the RNase H project and whose data chapters I've studied in some detail.

Chapter 1

Background

1.1 Relationships between the biophysical and biochemical properties of enzymes

1.1.1 Protein dynamics and stability

Structural changes are critical to the ability of proteins to execute biological function. This is particularly true of proteins that function as enzymes, due to the need to precisely position catalytic amino acid residues relative to one or more substrate or cofactor molecules, which must be bound stably enough to effect chemical catalysis but not so tightly that product molecules cannot be released back into the environment. However, the conformational plasticity needed for rapid catalytic turnover is necessarily in tension with the need for globular proteins to maintain a stable, compact folded structure. In the case of thermostability in particular, increased rigidity has long been recognized as a mechanism for tolerating the increased thermal fluctuations experienced by a protein at high temperature [1]. Therefore flexibility within a globular protein may trade off against resistance to thermal denaturation.

Mechanisms of protein stabilization have been extensively studied through the comparison of homologous proteins from organisms adapted to varying properties of the bulk environment. For this purpose, proteins found in organisms identified as “extremophiles”—that is, capable of surviving and thriving at extremes of temperature, pressure, environmental radiation, acidity, salt concentration, atmospheric chemical composition, and a variety of other environmental features—provide a fruitful platform for the investigation of adaptive changes in protein stability as a function of environmental pressures. Unsurprisingly, within a protein family, thermostability tends to correlate with the optimal growth temperature of the source organism [2]. The study of proteins from thermophilic organisms (and to a lesser extent their psychrophilic, or cold-adapted, cousins) has led to the identification of a number of general strategies for protein thermostabilization. Features thought to contribute to protein thermostabilization include more salt bridges [3], shorter loops [4; 5], better hydrophobic packing [3; 6], and global optimization in charge positioning [7] in proteins from thermophilic organisms compared to their homologs from mesophilic organisms. Comparisons of large sets of proteins from thermophilic and mesophilic bacteria of the same genus find that the thermophilic proteins are enriched in charged residues and depleted in uncharged polar residues [8]. Increased content of intracellular disulfide bridges has also been reported for some thermophilic prokaryotes [9]. Although relatively little work has been done on the thermal adaptations of eukaryotes, broadly similar patterns have been reported in the genome of a highly thermophilic eukaryote, the polychaete worm *A. pompejana* [10].

The advent of large-scale genomic sequencing projects has greatly facilitated the study of thermal adaptation, particularly among prokaryotes. A distinction has emerged between two broad categories of thermostabilization of proteins: a “structure-based” method in which thermostable proteins are significantly more compact than their mesophilic counterparts, whereby stabilization is contributed by an increase in total number of inter-residue interactions rather than by specific changes in the nature of these interactions, and which tends to occur in organisms that originated in hot environments; and a “sequence-based” method in which stabilization is achieved by introducing specific, strong interactions between individual, identifiable residues, which tends to occur in organisms that likely have mesophilic ancestors and later recolonized hot environments [11]. Recently, differences in the mechanisms of thermostabilization of archaeal and bacterial proteins have been described, with the former favoring improved hydrophobic packing and the latter favoring increased numbers of ion pairs, reflecting the structure- and sequence- based adaptive pathways, respectively [12]. A large-scale study of over 500 prokaryotic genomes observed increases in charged residues among the whole proteomes of both thermophilic archaea and bacteria [13].

The molecular mechanisms underlying the cold-tolerance of proteins from psychrophilic organisms have been much less well studied than their heat-tolerant counterparts. However, the patterns that have been observed are typically the reverse of those contributing to thermostabilization. Proteins from psychrophilic organisms are both relatively flexible and relatively thermolabile compared to their mesophilic homologs [14]. Whereas thermophilic proteins are enriched in charged residues, psychrophilic proteins are depleted [15]. As judged by both structural bioinformatics and whole-genome analyses of recently sequenced psychrophiles, proteins from these organisms also tend to feature longer loops, weaker hydrophobic packing, more solvent-exposed hydrophobic groups, and fewer salt bridges relative to mesophilic homologs [16].

A number of avenues of experimental evidence support the interpretation that increased rigidification of a protein correspondingly increases its thermostability. Increasing thermostability is associated with decreasing susceptibility to proteolysis, which is known to initiate at flexible sites [17; 18], and with decreasing hydrogen-deuterium exchange, which is a phenomenon that indicates exposure of core residues to solvent due to transient unfolding

[19; 20; 21]. Furthermore, successful approaches to rational design of thermostabilizing mutations typically emphasize the introduction of rigidifying features such as disulfide bonds, salt bridges, proline residues in loop regions, or optimized helix-capping interactions [22]. The seemingly naive approach of selective substitution of residues exhibiting high crystallographic B-factors (which indicate high positional uncertainty in the structural model) yields increased protein thermostability [23].

At the level of individual proteins, the separate thermodynamic contributions to protein stability can be measured and fit to protein stability curves described by the Gibbs-Helmholtz equation, which defines the temperature dependence of the free energy of unfolding [24]:

$$\Delta G(T) = \Delta H_m(1 - T/T_m) - \Delta C_p[T_m - T(1 - ln(T/T_m))] \tag{1.1}$$

where T_m is the melting temperature (that is, the temperature at which $\Delta G = 0$ and the protein is half folded and half denatured), ΔH_m is the change in enthalpy upon denaturation, and ΔC_p is the change in heat capacity upon denaturation. (Note that the derivation of this description makes the assumption that ΔC_p is constant over the temperature range in question.) Proteins have $\Delta C_p > 0$ [25]—that is, they experience an increase in heat capacity in the unfolded state—and experimental protein stability curves resemble Figure 1.1, where the zero-crossings of the parabola indicate the cold- and heat-denaturation temperatures T_c and T_m, and the maximum of the curve indicates the temperature of maximal stability T^*.

This equation implies that there are three primary mechanisms by which a protein can become more thermostable (that is, increase its T_m): (1) the curve can be shifted up, via an increase in ΔH_m); (2) the curve can be broadened, via a decrease in ΔC_p (which implies an increase in residual structure of the unfolded state); and (3) the curve can be shifted to the

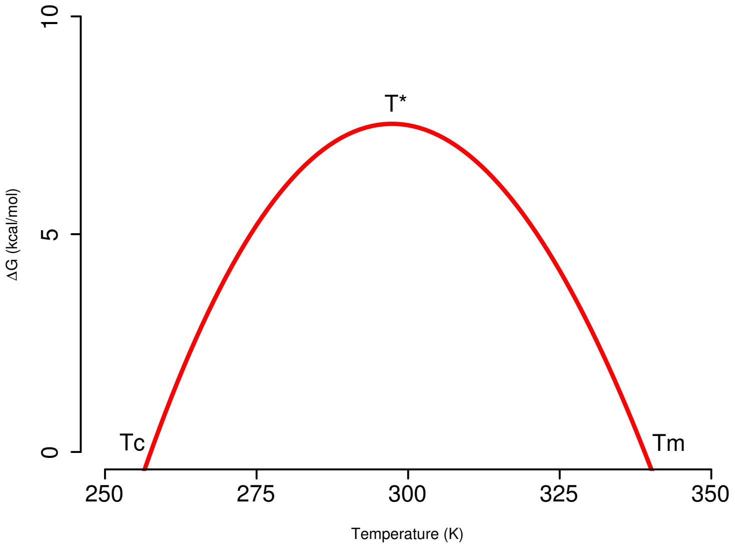

Figure 1.1: **An example protein stability curve.**
An example of a protein stability curve calculated according to Equation 1.1. The curve defines the temperature range over which $\Delta G > 0$, that is, where more than 50% of the protein molecules in a solution are considered to occupy the native state. The temperatures of cold denaturation T_s, heat denaturation (melting temperature) T_m, and maximal stability T^* are indicated.

right on the x-axis, via a decrease in ΔS_m (the change in entropy upon denaturation) [26; 27]. Examples of all three primary mechanisms of thermostabilization, as well as mixed modes featuring more than one of these mechanisms, have been identified in the comparisons of mesophilic and thermophilic proteins. Intriguingly, the most commonly observed mechanism of thermostabilization is a combination of up-shifting and broadening—that is, increases in ΔH_m and decreases in ΔC_p—that results in a conserved value of T^* within a family [27]. Furthermore, the tendency of T^* values to cluster around room temperature—that is, "mesophilic" temperature—has been observed in formal exploration of the parameter space of the Gibbs-Helmholtz equation[28]. Relatedly, the value of ΔG for a particular protein family tends to be comparable at the optimal growth temperature of each protein's source organism [29; 30]. This observation can be interpreted as evolutionary conservation of the balance between dynamics and stability, which leads to the hypothesis that specifically functional aspects of protein dynamics can be isolated through the comparative analysis of members of protein families from organisms adapted to different thermal environments.

1.1.2 Protein dynamics and catalytic activity

The role of protein dynamics in relating thermostability to catalytic activity has been a subject of significant interest, not only as a matter of basic research but also as a source of potential applications in the biotechnology industry. Enzymes are highly efficient catalysts and have long held promise as environmentally benign and cost-effective tools for chemical manufacturing, food processing, and bioremediation. The rational design of proteins that are highly active in extreme environments and nevertheless highly stable offers a variety of potential industrial applications [31; 32; 33].

The detailed relationship between protein dynamics and catalytic activity has nevertheless remained difficult to characterize experimentally. Nuclear magnetic resonance (NMR) spectroscopy has emerged as the major experimental tool for studying protein dynamics at atomistic resolution and has been exceptionally productive in identifying protein features such as regions of high flexibility over a variety of timescales [34; 35; 36]. Regions known to be in contact with substrates and to undergo conformational changes during the catalytic cycles of enzymes often are identifiable as particularly flexible by NMR spectroscopy [35; 37] and by computational methods [38; 39; 40]; however, mechanistic descriptions of the structural changes underlying flexibility are difficult to establish. Molecular dynamics (MD) simulations can complement observations made by NMR via direct simulation of functionally relevant dynamic processes [41; 42; 43; 44].

The relationship between conformational dynamics and catalysis has been the subject of extensive recent debate [45; 46; 47]. Although the majority of the controversy has focused on the question of whether dynamics have an effect on the chemical step in the catalytic cycle—and at best, the effect seems to be limited to fast-timescale, local motions, possibly restricted to those cases in which enzymes' catalytic mechanisms rely on hydrogen tunneling [48; 49]—questions remain regarding the role of dynamics in binding and orienting substrate and cofactors to generate the precise electrostatic preorganization thought to be required for catalysis [50; 51]. Thus larger-scale motions of enzymes, particularly in those regions known to interact with substrate, influence binding affinity, product release rates, and other processes relevant to determining the overall function of the enzyme.

Homologous sets of proteins derived from organisms adapted to different thermal environments have proven especially useful in understanding the functional aspects of protein dynamics [52; 53; 54; 55; 56]. A number of cases have been identified in which a thermophilic enzyme is both more rigid and less active than its mesophilic homolog at ambient temperature [53; 21; 57]. One well-characterized example comes from the adenylate kinase enzyme family, in which the opening of a substrate-binding lid occurs at a rate commensurate with the overall cataytic rate, and is significantly slower at ambient temperature in a hyperthermophilic homolog relative to its mesophilic counterparts, thereby straightforwardly explaining the reduced activity of the thermostable protein [53]. Furthermore, *in vivo* laboratory evolution of thermotolerance is directly attributed to an adenylate kinase point mutant that confers both increased rigidity and a significant increase in enzymatic activity at high temperature, at the direct expense of activity at lower temperature [58].

Such observations have led to the the hypothesis that motions critical to function can be specifically identified by comparing the dynamics of homologs from organisms with different preferred growth temperatures. A hypothesis widely used as an interpretive framework in the study of thermal adaptation defines the notion of "corresponding states" [59], according to which homologous proteins should have similar degrees of structural flexibility (and occupy similar overall conformational ensembles) in the optimal temperature ranges of their source organisms. In order to achieve this matched flexibility, proteins from thermophilic organisms might be expected to exhibit reduced flexibility at temperatures below those preferred by their source organism [21]. Alternatively, they might exhibit similar flexibilities over a wider temperature range than their mesophilic homologs; although the structural mechanisms by which this could occur are less clear, such cases have been described [60].

In some well-studied protein families, however, the corresponding-states hypothesis does not map well onto observations that have been made of the relative flexibilities of homologous proteins. Perhaps the clearest example of such deviation comes from the α-amylase family, in which increased flexibility is noted in the homolog with higher melting temperature [61; 62; 63]. Thermophilic rubredoxins provide another example, in which thermostability is attributed to differences in unfolding rates [64] and in the dependence of flexibility on temperature [60]. Studies on dihydrofolate reductase are numerous and often contradictory, but both experimental [65] and theoretical [66] work has suggested that the thermophilic homolog is in fact more flexible.

The ribonuclease HI (RNase H) homologs from the mesophilic bacterium *Escherichia coli* (ecRNH) and the thermophilic bacterium *Thermus thermophilus* (ttRNH) comprise an extremely well-studied homologous pair [67; 68; 69; 70; 71] which have been reported not to fit the corresponding-states model as determined by experiment [72; 73] and simulation [74]. Since these studies were conducted, several additional members of the family from organisms of varying growth temperature have been structurally characterized, affording an opportunity to explore the thermal adaptation of conformational dynamics in this family in much greater detail. Interestingly, a relationship has been described between the conformational dynamics of individual proteins and the molecular diversity of their evolutionary homologs [75], further motivating the comparative study of additional members of the RNase H family.

1.2 Ribonucleases HI

1.2.1 Properties of the RNase H superfamily

RNase H proteins are well-conserved endonucleases that are found in all domains of life and sequence-agnostically cleave the RNA strand of an RNA-DNA duplex substrate in a divalent cation-dependent manner, producing a free 3' OH group [67]. RNase HI (the protein corresponding to the product of the *rnhA* gene) is the founding member of the superfamily and is the subject of the present work. The RNases HII and HIII families (*rnhB* and *rnhC* respectively) are closely structurally related to each other, although much more distantly related to RNase HI in overall fold [76]; the structural superposition between *Homo sapiens* RNase H1 and H2 proteins is illustrated in Figure 1.2 (note that eukaryotic nomenclature for these proteins dictates Latin rather than Roman numerals [77]). All three protein families nevertheless share a common canonical active-site organization consisting of a DED(D) motif, three to four carboxylate-containing residues collectively participating in cation binding and likely in acid-base chemistry during enzymatic catalysis [67]. This fundamental pattern is widely shared with other nucleases and polynucleotidyl transferases involved in a variety of other biological processes, including RuvC Holliday junction resolvase, the PIWI domain of Argonaute proteins involved in RNA interference and related processes, as well as other retroviral activities such as integrase and transposase [67].

Although a wide variety of conserved biochemical processes demand RNase H enzymatic activity, the functional specialization of these three families is not well understood and appears to vary among organisms. In prokaryotes, RNase HI is not essential; however, it has been implicated in the suppression of chromosomal DNA replication from locations other than the canonical *oriC* initiation site [78]; plays a secondary role to DNA polymerase I in removal of lagging-strand RNA primers during Okazaki fragment processing in DNA replication [79]; and participates in the removal of transcriptional byproducts known as

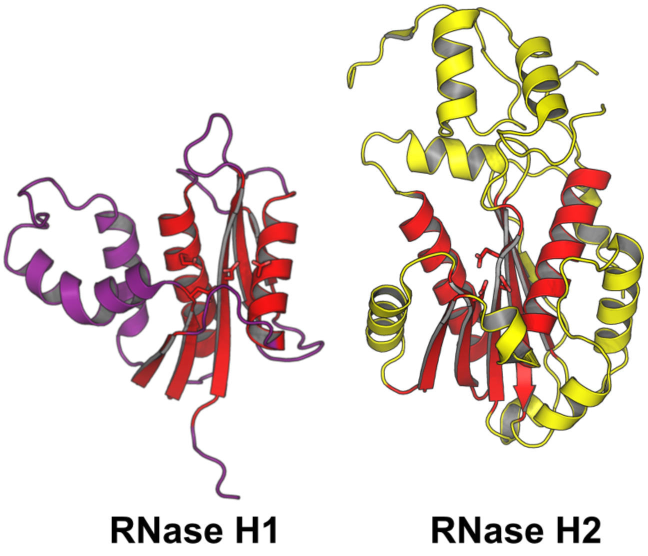

Figure 1.2: **Structures of *Homo sapiens* RNases H1 and H2.** RNases H1 (right, PDB ID 2QK9) and H2 (left, PDB ID 3PUF) have different overall folds but nevertheless share similar core architecture and active-site organization. Structurally superposed regions are shown in red for each protein. Active-site residues are shown as sticks.

R-loops [80]. Interestingly, although the formation of R-loops was once considered to be a rare transcriptional error, they have recently been identified as common sources of genomic instability [81], the prevention of which specifically requires RNase H1 in both yeast [82] and humans [83].

The *H. sapiens* genome contains homologs of *rnhAB* (denoted RNASEH1 and RNASEH2 in human genome nomenclature), although in human and many other eukaryotes the RNase H2 protein has become an obligate heterotrimer [77]. Although mutations in any of the three human RNase H2 subunits are associated with a genetic disease known as Aicardi-Goutières syndrome, which affects neurological function and is fatal in early childhood, no known disease-related mutations have been identified in RNase H1 [77]. Known single-nucleotide polymorphisms that give rise to nonsynonymous mutations are summarized in Figure 1.3; they are distributed widely on the surface of the protein but do not appear in the core [84] (Ensembl v.72). RNase H1 function is likely essential in higher eukaryotes. Although yeast deletion strains have been successfully produced [85] with only mild deficiencies in stress response [86], loss of RNase H function is lethal in both flies and mice. In *Drosophila* knockout mutants cell proliferation *per se* is not deficient, but the mutation is nevertheless lethal at the pupation stage, suggesting that RNase H is required for metamorphosis [87]. Mouse models lacking the RNASEH1 gene die in early embryogenesis due to defects in mitochondrial genome processing [88], a function that seems likely to parallel the role played in human development, since the human gene product includes a mitochondrial targeting sequence [77] and is required for Okazaki fragment processing during mitochondrial DNA replication [89].

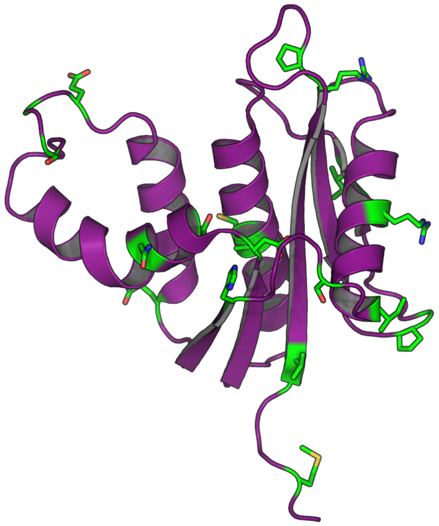

Figure 1.3: **Known non-synonymous substitutions in RNase H1 in the human genome.**
Sites at which a non-synonymous substitution has been reported are highlighted in green on the structure of the human RNase H1.

In prokaryotes, RNase HI often appears as a single domain [67]. In eukaryotes it is most typically the C-terminal domain of a two-domain protein containing an N-terminal hybrid binding domain (HBD) attached by a long flexible linker [77]. This domain binds nucleic acids but has no enzymatic function; it likely serves as a processivity factor[90; 91].

Additionally, RNase HI domains appear as a component of the retroviral reverse transcriptase complex, in which they are required for viral proliferation [92]. This observation has focused interest on the RNase HI family as a possible drug target for antiretroviral medications as well as a model system for the study of protein dynamics.

In the remainder of the present work, the term “RNase H” will be used for simplicity to refer to the RNase HI family unless otherwise specified.

1.2.2 Molecular diversity among structurally characterized RNases H

Among the RNase H family members, by far the best characterized is the homolog from *Escherichia coli* (ecRNH). The structure of this 155-residue protein, solved independently by two research groups [93; 94; 95], represented at the time a novel fold now known as the RNase H fold. The structure consists of a β-sheet of mixed parallel and antiparallel strands, flanked by helices packed on both surfaces of the central sheet. A prominent feature of the ecRNH structure is the region referred to as the handle region, also known as the basic protrusion due to its high density of positive charge; this region is absent in many distant homologs of RNase H, but is shared among the four structurally characterized bacterial proteins. Key features of the structure are illustrated in Figure 1.4.

The ecRNH structure has been extensively studied in comparison to its thermophilic homolog from *Thermus thermophilus* (ttRNH); despite 55% sequence identity with ecRNH and less than 1Å C_α RMSD in secondary structural elements, ttRNH has reduced enzymatic activity [68] and greater thermal stability [69; 70] compared to ecRNH. Reciprocal mutations have identified five distinct sites that collectively contribute about half of this stability

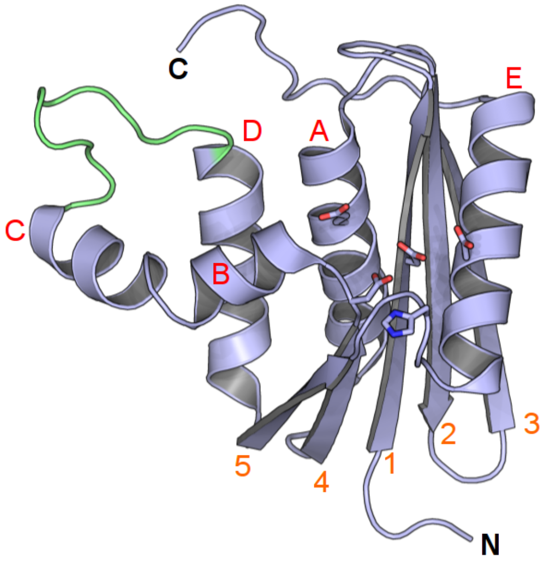

Figure 1.4: **Key features of the ecRNH structure.**
The ecRNH structure (PDB ID 2RN2) is shown with its β sheets labeled in orange and α helices labeled in red. Active-site residues are shown as sticks. The handle loop is highlighted in green.

difference [96; 71]. More recently, similar analyses have identified mutations that confer increased thermostability to the homolog from the psychrotrophic bacterium *Shewanella oneidensis* (soRNH) [97]; like many proteins from cold-tolerant organisms [14], soRNH is natively thermolabile compared to its mesophilic homolog. The largest single-residue contribution to the thermostability of ttRNH is the identity of the residue at position 95 [98], which is conserved in a left-handed region of Ramachandran space and is a glycine in ttRNH, compared to lysine in ecRNH and arginine in soRNH. Sites known to make significant contributions to thermostabilization in these three proteins are illustrated in Figure 1.5.

Comparison of the thermodynamic parameters of these proteins, along with an additional homolog from the moderately thermophilic bacterium *Chlorobium tepidum*, reveals that the more thermostable proteins share a common mechanism of stabilization in the

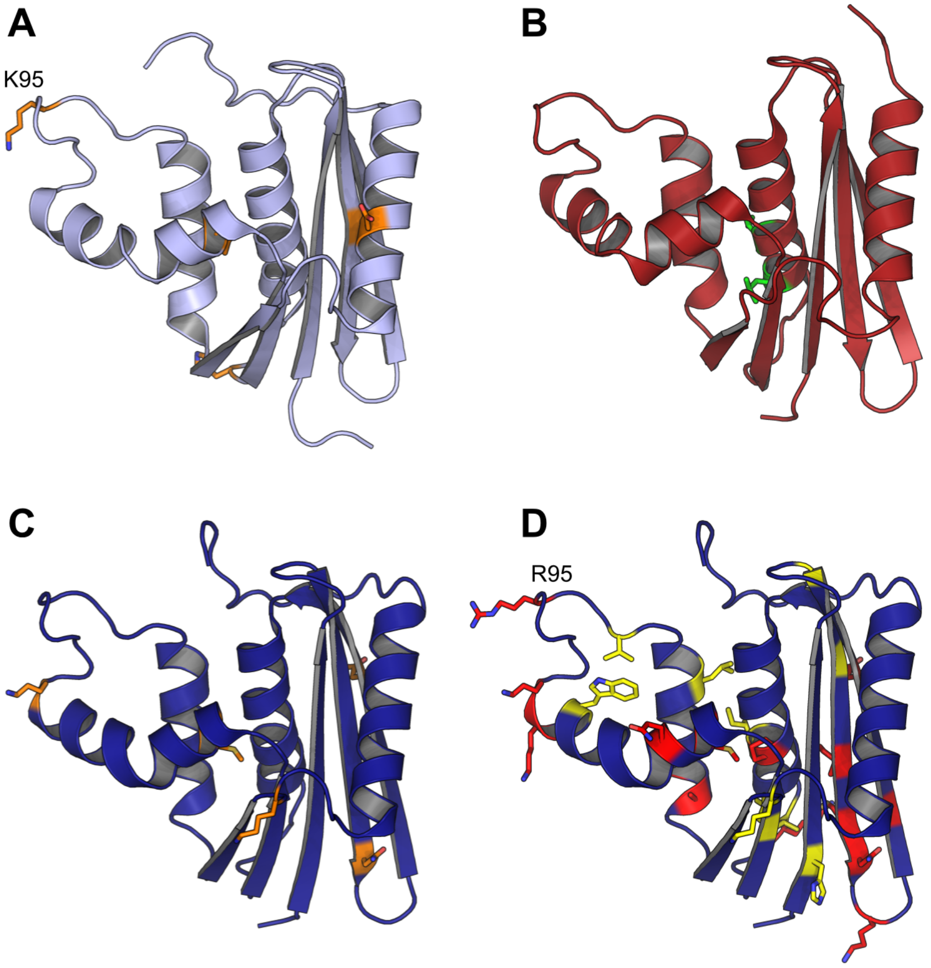

Figure 1.5: **Known sites of thermostabilization in the RNase H family.** (A) Sites identified as increasing the stability of ecRNH [96; 71]. (B) Sites identified as contributing to residual structure in the unfolded states of ttRNH and ctRNH [70; 99]. (C) Sites identified as increasing the stability of soRNH [97]. (D) Results of a large-scale mutagenesis study designed to identify stabilizing mutants of soRNH. Sites shown in red indicate that at least one mutation was observed that increased the stability of the protein, while sites shown in yellow indicate that no mutation was found to confer stability at that site [100]. The position of residue 95, established as a locus of thermostability in ttRNH [98], is highlighted; two of the three stabilization screens identified its substitution as a source of thermostability.

form of decreased values of ΔC_p, likely owing to the existence of residual structure in the unfolded state [101; 99]. Figure 1.6 shows the stability curves for the four RNase H homologs whose thermodynamic properties have been characterized; the corresponding values are summarized in Table 1.1.

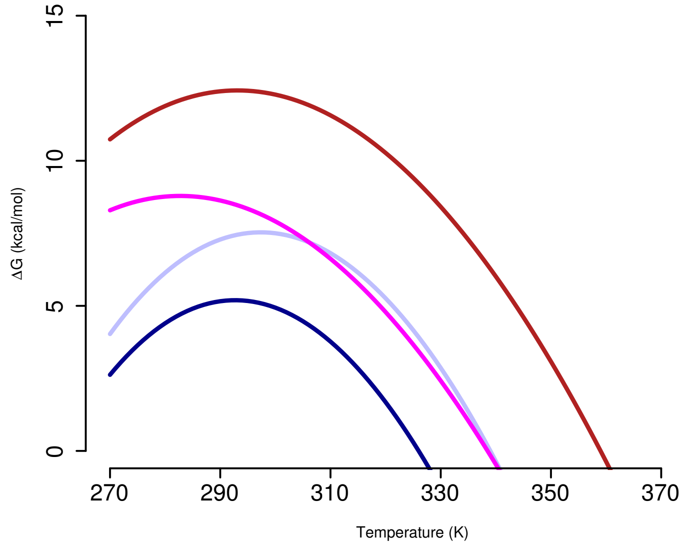

Figure 1.6: **Protein stability curves for the four RNase H homologs whose properties have been studied.**
Stability curves are shown for soRNH (dark blue), ecRNH (light blue), ctRNH (magenta), and ttRNH (red). Data is reproduced from [30] for ecRNH and ttRNH, [102] for soRNH, and [101] for ctRNH. As summarized in Table 1.1, ctRNH exhibits a decreased ΔC_p (that is, a broadened curve) relative to ecRNH without change in other parameters, particularly T_m, while ttRNH exhibits both broadening and up-shifting. Relative to soRNH, ecRNH is right- and up-shifted, a pattern consistent with observations of other psychrophile-mesophile pairs [63]. Thus all three modes of stabilization are represented within this protein family. Note that values for ecRNH, ctRNH, and ttRNH are reported for the cysteine-free variants, which tend to be approximately $3°C$ destabilized in T_m relative to wild type.

1.2.3 Folding and site-specific stability in RNases H

The RNase H family has long been used as a model system for the study of protein folding processes in globular proteins. Extensive study of the folding process of ecRNH has

Table 1.1: **Thermodynamic properties of RNase H homologs.**

Protein	ΔH_m (kcal/mol)	ΔC_p (kcal/mol $\cdot$ K)	T_m (K)
soRNH	99.9±3.6	2.8±0.3	326.2±0.5
ecRNH	120±4	2.7±0.2	339±1
ctRNH[a]	103	1.7	338.5
ttRNH	131±5	1.8±0.1	359±1

Thermodynamic properties of the four RNase H homologs that have been studied, corresponding to the values used to construct the protein stability curves in Figure 1.6. [a], errors were not reported for ctRNH, and ΔH_m was back-calculated from the reported ΔG_{unf} value. Data is reproduced from [30] for ecRNH and ttRNH, [102] for soRNH, and [101] for ctRNH. Note that values for ecRNH, ctRNH, and ttRNH are reported for the cysteine-free variants, which tend to be approximately $3°C$ destabilized in T_m relative to wild type.

revealed that it forms its native structure via a three-state folding mechanism, with a single marginally stable kinetic intermediate [103; 104]. This intermediate structure closely resembles the partially denatured states accessible under conditions such as acid denaturation [105; 106] and mechanical unfolding [104], as well as a molten globule-like state that exists in equilibrium with the protein's native state [107; 108; 106]. Characterization of the interactions present in this state by ϕ-value analysis reveal that the folding core of the protein significantly overlaps with the hydrophobic core present in the folded state [103] and localizes to the dimeric coiled-coil structure formed by helices A and D, as well as β-strand 4 [106]. Specific mutations in which this folding core is disrupted due to the introduction of a charged residue result in destabilization of the folding intermediate to produce a two-state folding mechanism and reduced native-state stability [109] Monitoring both the native-state equilibrium with partially denatured forms [110] and the folding process of the wild-type protein identifies hierarchical folding and unfolding processes in regions of the protein that can be interpreted as distinct folding units, or "foldons" [111] (Figure 1.7).

The ecRNH and ttRNH native-state stabilities and folding processes have been directly compared in order to elucidate the origins of ttRNH's higher thermostability. Examin-

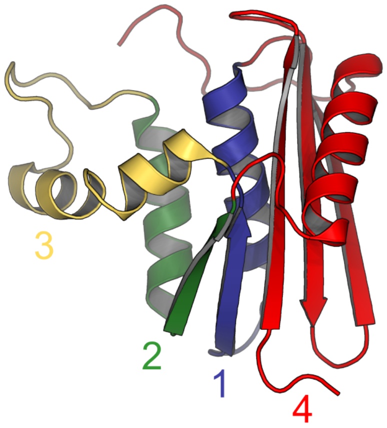

Figure 1.7: **Hierarchical folding regions in ecRNH**
The five distinct "foldons" identified in ecRNH are shown with the order of folding indicated (data and coloring from [111]).

ing the denaturation equilibria of the two proteins suggests that the increased stability of ttRNH is delocalized across the entire protein, whereby nearly all of the secondary structural elements whose stabilities can be measured are more stable in ttRNH [110]. Single-site thermostabilizing mutations in ecRNH do not recapitulate the ttRNH pattern, underscoring the interpretation that the balance of thermostability and flexibility in ttRNH is evolutionarily adaptive [112]. The process of folding, including the presence of a kinetic folding intermediate, is well-conserved in ttRNH [70]. The contributions of two proteins' folding cores to thermostability can be compared by construction of chimeric proteins in which the folding core residues are reciprocally swapped between the two homologs; this study reveals that thermostability is largely conferred by the presence of the ttRNH folding core [113]. Additionally, mutations disruptive to the ecRNH folding intermediate also destabilize ttRNH, specifically by perturbing its unusually low ΔC_p value, interpreted as disrupting residual structure present in the ttRNH unfolded state [114] (Figure 1.5B).. Interestingly, similar mutations in the folding core of ctRNH suggest that although it too has a low ΔC_p value, its residual unfolded structure is distinct from that of ttRNH and is sensitive to mutations at distinct sites [99].

The folding processes of RNases H have also been studied in the presence of Mg^{2+} ions, which are known to increase the thermostability of ecRNH due to reduction of electrostatic repulsion in the active site [115]. In ecRNH, the folding process is not significantly altered by the presence of Mg^{2+}, implying that metal binding occurs only after folding is complete [116]. However, contributions of Mg^{2+} ions to the folding process have been observed in a retroviral RNase H domain [117], suggesting distinct structural origins of native-state stability within bacterial and retroviral RNases H.

1.2.4 Substrate interactions and chemical mechanism

To date, the only RNase H homolog whose structure has been solved in both the presence and absence of substrate is the homolog from *Bacillus halodurans* [118], which lacks the handle region and therefore is a poor model for understanding the behavior of the ecRNH protein. The structure of the *H. sapiens* (hsRNH) homolog has been solved in the presence of substrate and has extremely similar three-dimensional structure [119], making it the best available model for the bacterial proteins. Comparison of the ecRNH and hsRNH structures (Figure 1.8) reveals three loops that change conformation upon interaction with substrate: the glycine-rich loop located between $\beta 1$ and $\beta 2$, the handle loop located between αC and αD, and the active-site loop located between $\beta 5$ and αE.

Examination of the hsRNH-substrate complex reveals that the scissile phosphate group of the RNA strand is located between two coordinated metal ions in the active site, and that specific interactions with the DNA backbone are formed in two distinct positions in the substrate-binding interface—a “phosphate-binding pocket” located near the center of the interface, and a DNA-binding channel formed by the handle loop—that would sterically exclude RNA [119].

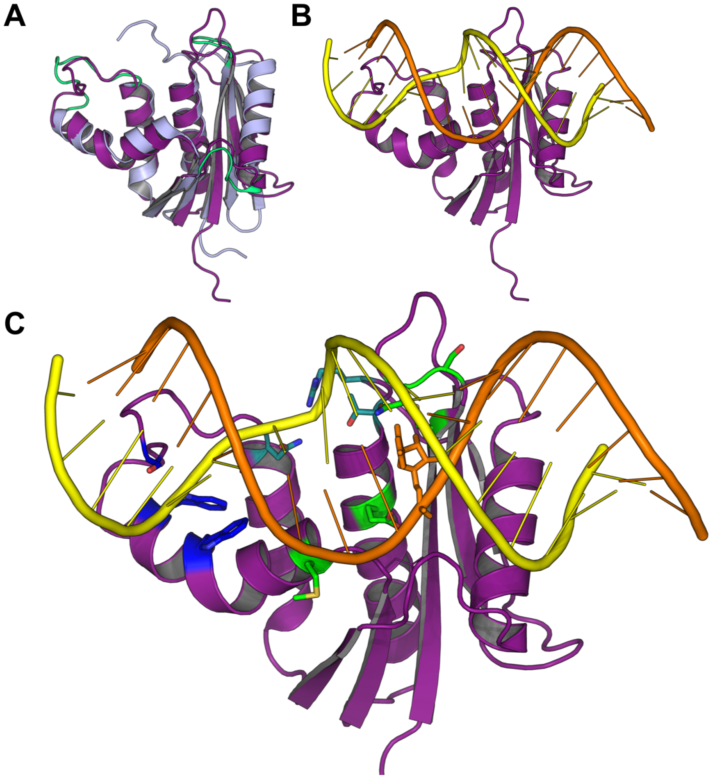

Figure 1.8: **Mechanisms of RNase H interactions with substrate.**
(A) The ecRNH protein (light blue) superposed on the hsRNH protein (purple). Flexible loops that change conformation upon substrate binding are shown in green. (B) The hsRNH-substrate complex. DNA is shown in yellow; RNA is shown in orange. (C) Three groups of residues identified as contributing to substrate interactions and specificity by inspection of the hsRNH complex [119]. Residues in green contribute binding to the backbone of the RNA strand, residues in cyan form the phosphate-binding pocket specific to deoxyribose backbones, and residues in blue form the DNA-binding channel formed by the handle loop and helix C. The base contributing the scissile phosphate is shown as sticks.

The exact chemical mechanism through which RNase H performs its function of cleaving the RNA strand has historically been controversial. Catalysis is dependent on the presence of divalent cations, physiologically Mg^{2+}, though other ions, particularly Mn^{2+}, have also been reported to support enzymatic activity. Cocrystallization of ecRNH with Mg^{2+} found only a single ion bound in the active site [120]; however, crystallization in the presence of Mn^{2+} yields two bound ions [121]. Chemical mechanisms requiring one [122; 123; 124] and two metal ions [125; 118; 126] (the latter by analogy to the Steitz mechanism for catalytic RNA [127]) have been proposed. Quantum mechanics/molecular mechanics (QM/MM) computational approaches tend to support the two-metal mechanism [128; 129].

Enzymatic activity has been studied in detail for three RNase H homologs: soRNH, ecRNH, and ttRNH. Available kinetic data is summarized in Table 1.2. The pattern observed for ecRNH and ttRNH parallels those observed for other mesophile-thermophile pairs: ttRNH is significantly less active at ambient temperature compared to ecRNH, but is extremely active closer to the optimal growth temperature of its source organism. Similarly, soRNH is less active than ecRNH at ambient temperature, as has been reported for other cold-adapted proteins [130].

Table 1.2: **Available kinetic measurements for RNase H homologs**

Protein	Temp ($°C$)	Substrate	K_m (μM)	k_{cat} (min^{-1})	V_{max} (units/mg)	Reference
soRNH	30	M13	0.30	–	8.6	[102]
ecRNH	30	M13	0.11	–	9.5	[102]
ecRNH	30	9-mer	0.53	90	–	[68]
ttRNH	30	9-mer	3.9	19	–	[68]
ecRNH	37	M13	0.11	–	36	[68]
ttRNH	37	M13	0.5	–	7.5	[68]
ttRNH	70	M13	1.1	–	104	[68]

Kinetics data measured under various conditions for soRNH, ecRNH, and ttRNH.

1.2.5 Dynamics in the RNase H family

As one of the earliest examples of a mesophile-thermophile pair whose members were both structurally characterized, the conformational dynamics of ecRNH and ttRNH have also been the subject of extensive study by NMR [131; 132; 133; 134; 135; 72; 73]. Superposition of the two structures reveals no major conformational differences (Figure 1.9), leading to the hypothesis that the origin of their large differences in activity might lie in different dynamic properties.

Only relatively subtle differences in dynamic behavior between the two proteins were observed on the ps-ns timescale [72]: dynamics in ttRNH do not indicate significantly higher global rigidity; however, particular regions of the protein—most notably, the two β strands at the exterior edges of the central β-sheet—are more rigid in ttRNH, possibly contributing to its relative thermostability. Interestingly, the handle loop is slightly but significantly more flexible in ttRNH. Dynamics on the ps-ns and μs-ms timescales are summarized in in Figure 1.9.

The most prominent difference in sequence is the presence of an inserted glycine residue in the junction between helices αB and αC in ttRNH; this inserted residue is conserved in ctRNH, the other structurally characterized example of an RNase H homolog from a thermophilic organism. NMR studies of wild-type ecRNH and ttRNH reveal the presence of conformational dynamics at the μs-ms timescale in this region in ttRNH that are not present in ecRNH [73]. Moreover, the rate of this dynamic process is closely matched with the rate of dynamics observed in the active-site loop, suggesting that motions in the handle and near the active site may be dynamically coupled.

The insertion of this glycine residue into the ecRNH sequence (denoted ecRNH iG80b) does not significantly affect its thermostability and produces only subtle changes in its structure [136]. However, this mutation and its reciprocal deletion mutant from ttRNH (denoted ttRNH dG80) does significantly affect both catalytic activity and dynamics. The ecRNH iG80b mutation significantly reduces activity [136] and confers μs-ms dynamic behavior that resembles wild-type ttRNH [73]. Conversely, ttRNH dG80 exhibits dynamic behavior that resembles wild-type ecRNH—that is, an absence of μs-ms dynamics in the

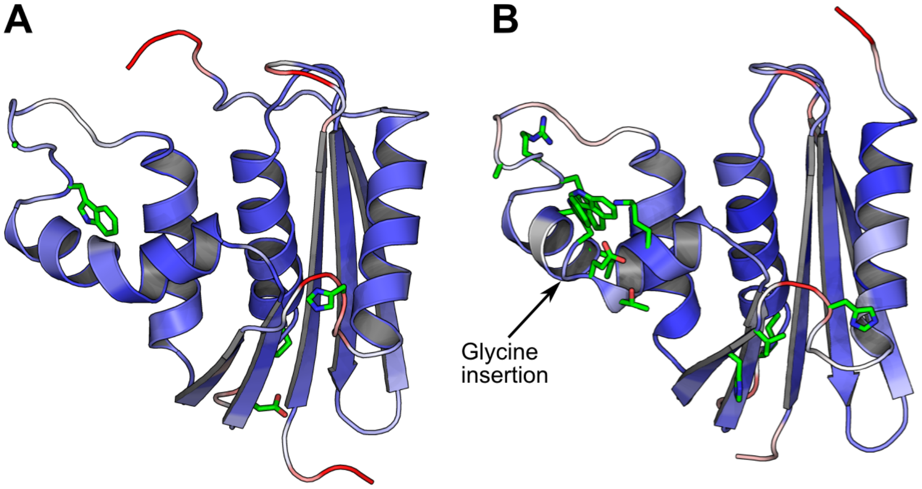

Figure 1.9: **Dynamics of ecRNH and ttRNH on the ps-ns and μs-ms timescales.** Backbone amide S^2 values indicating motion on the ps-ns timescale are mapped onto the structures of (A) ecRNH, and (B) ttRNH, with blue indicating rigidity and red indicating flexibility. For both proteins, residues identified as experiencing chemical exchange line broadening, which indicates dynamics at the μs-ms timescale, are shown as sticks in green.

helix junction [73]—although without increase in catalytic activity (J.A. Butterwick, unpublished data).

Despite this history of extensive investigation, the relationships between dynamics, thermostability, and enzymatic activity in the RNase H family remain obscure. The purpose of the present work is to exploit the complementary nature of molecular dynamics simulations and NMR spectroscopy to provide atomistic interpretations of the conformational dynamics conserved among the RNase H family. Properties of the family members whose dynamics were studied in the present work are summarized in Table 1.3. The results presented herein illustrate the utility of combined MD-NMR studies in understanding molecular adaptation at the level of individual residues to features of the bulk environment.

Table 1.3: **Properties of structurally characterized RNase H homologs.**

Protein	Source organism	Classification	C_α RMSD (Å)	% Seq ID
soRNH	*Shewanella oneidensis*	Psychrotroph	1.37	70%
ecRNH	*Escherichia coli*	Mesophile	–	–
ctRNH	*Chlorobium tepidum*	Thermophile	1.95	49%
ttRNH	*Thermus thermophilus*	Thermophile	1.44	55%
hsRNH	*Homo sapiens*	Mesophile	2.53	34%

Properties and source organisms of the non-retroviral RNase H homologs studied in this work. C_α RMSD and percent sequence identity are relative to ecRNH.

1.3 Molecular dynamics simulations

1.3.1 Foundational principles

All-atom molecular dynamics (MD) simulations are the only current computational tool capable of providing the high spatial and temporal resolution necessary for atomistic interpretations of the motions of biological macromolecules. The basic principle underlying MD simulations is a very simple one: Newton's equations of motion are numerically integrated over time for a large but finite number of discrete interacting particles (typically representing a condensed-phase system), producing a trajectory that describes the internal dynamics of the system as a function of time. Although the method originated in theoretical physics [137] and was first applied primarily to simple systems such as liquid argon [138], modern uses of the method are primarily concentrated in polymer physics and biophysics. Of course, it must be emphasized that this is a classical treatment; quantum effects are not modeled beyond the extent to which they can be averaged or approximated by parameterization within the classical paradigm. To carry out an MD simulation, time is discretized into a succession of very short timesteps Δt and forces and particle positions recalculated at each step; an extremely simplified description of an MD algorithm follows:

1. Define a set of initial positions $\vec{r}^{t=0}$ for the set of vectors $\vec{r}^N = \vec{r}_1, \vec{r}_2, ..., \vec{r}_N$ for N particles in the system
2. Assign initial velocities $\vec{v}^{t=0}$
3. Define an inter-particle interaction potential $V(\vec{r}^N)$
4. Calculate net force on each particle i, $\vec{F_i} = -\sum_{j=1}^{N} \nabla V_i(r_{ij}), i \neq j$
5. Update particle positions $\vec{r}^{t+\Delta t} = \vec{r}^t + \vec{v}^t \Delta t + \frac{1}{2}\vec{a}\Delta t^2 + ...$
6. Increment time $t = t + \Delta t$
7. Goto 4

In practice, this simple scheme of updating positions is not numerically stable; most commonly a two-step scheme such as the velocity Verlet method is used [139] and extensive theoretical work has been done on the development of improved methods for numerical integration (eg [140]). Furthermore, additional updates may be made in order to control system variables such as temperature and pressure [141; 142; 143; 144; 145]. An MD simulation without such controls is referred to as sampling the NVE or micro canonical ensemble, in which number of particles (N), system volume (V), and total energy (E) are conserved; other commonly sampled ensembles are the NVT or canonical ensemble (constant volume and temperature) and the NPT or isothermal-isobaric ensemble (constant pressure and temperature).

In typical biomolecular simulations, the positions of atoms are initially defined by an experimentally determined macromolecular structure, most commonly solved by X-ray crystallography, with the addition of an appropriately sized buffer of solvent molecules for those simulations in which solvent will be explicitly represented. The velocities are initialized using a random sample from the Boltzmann distribution at a temperature of biological interest. A biomolecular system typically contains many different types of interacting particles whose interaction potential must be defined. A set of parameters that collectively defines this potential is known as a force field. Modern force fields decompose the interaction potential into a sum of terms:

$$E_{tot} = E_{bonded} + E_{nonbonded} \tag{1.2}$$

$$E_{bonded} = E_{bonds} + E_{angles} + E_{torsions} \tag{1.3}$$

$$E_{bonded} = \sum_{bonds} (r - r_0)^2 + \sum_{angles} (\theta - \theta_0)^2 + \sum_{torsions} V_n (1 + \cos(n\omega - \gamma)) \tag{1.4}$$

$$E_{nonbonded} = E_{VDW} + E_{Coulomb} \tag{1.5}$$

$$E_{nonbonded} = \sum_{i<j} \epsilon_{ij} \left[(\sigma_{ij}/r_{ij})^{12} - (\sigma_{ij}/r_{ij})^{6} \right] + \sum_{i<j} q_i q_j / \left(4\pi\epsilon_0 r_{ij}\right) \tag{1.6}$$

In this formulation, the potential energy is decomposed into a sum of bonded terms, pertaining to atoms directly connected by a chemical bond, and nonbonded terms, pertaining to atoms that are neighbors in three-dimensional space. The bonded terms consist of harmonic potentials to define bond lengths r and angles θ, and Fourier series with coefficients V_n and phase γ to define dihedral torsions ω. The nonbonded terms consist of a Lennard-Jones interaction to represent the van der Waals forces and a Coulombic potential to represent electrostatic interactions between particles of fixed atomic charge. In traditional simulations on commodity hardware, these latter two nonbonded terms account for the overwhelming majority of the computational cost, and are typically truncated through the use of distance cutoffs beyond which the interaction is assumed to be zero; this is a much more problematic assumption for the electrostatics term [146]. Modern methods almost universally have adopted the particle-mesh Ewald summation method [147] under periodic boundary conditions for accounting for long-range electrostatics. Parameter sets are typically derived from high-level quantum-mechanical calculations, and increasingly often are modified based on empirical comparisons with NMR data to improve agreement with experiment [148; 149; 150; 151]

Several features of this decomposition are worthy of note as they pertain to biomolecular systems. First, the force field terms are written as sums of pairwise interactions. This means that some interactions with relatively large contributions from many-body terms will necessarily be poorly modeled; for biological systems, particles with high charge density, such as divalent cations, are most affected by this approximation and require careful parameterization [152]. Second, because the bonded terms are modeled with harmonic potentials,"bond-breaking" chemistry cannot be modeled with traditional molecular mechanics force fields. This means that the internal dynamics of biomolecular systems can be explored, but events such as enzymatic catalysis are outside the scope of this model. Third, the force field treats individual particles as having fixed partial atomic charges, which do not update depending on their local environment. Force fields incorporating polarizable electrostatics have long been under development, although relatively few contexts thus far have demonstrated clear superiority for these more computationally intensive methods [153; 154].

Constraints on the choice of Δt derive from the numerical stability of the integration scheme and the dynamics sampled in the system; one chooses the timestep such that sufficient sampling is expected of the fastest motions accessible to the system. For biomolecular systems, this typically implies a Δt of approximately 1fs; however, the fastest motions in such a system correspond to bond vibrations of bonds involving hydrogen atoms, and the development of algorithms to constrain these vibrations has allowed the use of timesteps of 2-2.5fs. Furthermore, the development of the reference system propagator algorithm (RESPA) allows the use of longer timesteps specifically for the slower dynamics in the system, typically the "far" nonbonded interactions [155], which can be safely updated as infrequently as 7.5fs.

The primary limitations on the utility of an MD simulation derive from two sources: inadequate sampling time, and poor force field parameters. The former problem can be understood by considering the ergodic hypothesis: the time-averaged properties of the system converge to the statistical-ensemble properties. In other words, given infinite running time with a perfect numerical integration scheme, the simulation should completely sample the available conformational space, weighted according to the Boltzmann distribution. However, limitations of hardware and resources limit the total simulation time in practice, with the result that the simulation may be too short to adequately sample conformations of interest. This problem is exacerbated by the possibility of initiating a trajectory in a region of conformational space that is not representative of a highly populated state in the "real" ensemble—this can occur due to poor quality of the initial structure [156], or occasionally due to the details of system preparation [157]. The latter problem, poor force field parameters, dictates that even with infinite sampling time and perfect numerical behavior, the simulation still might not produce representative dynamics with conformational populations that reflect those observed in the experiment, because the conformational landscape defined by the force field is not representative of the native landscape. In addition to difficulties in parameterizing the solute particles of the system (proteins, nucleic acids, etc.), force field limitations may also derive from parameterization of solvent; for example, most commonly used models for water have been found to deviate appreciably from experimental behavior, particularly in their self-diffusion constants and therefore their estimates of overall tumbling times for large solute molecules [158; 159]. In practice, it is often difficult to distinguish between limitations due to sampling and due to force field errors for a given simulation.

1.3.2 Evolving relationship to NMR

The first biomolecular dynamics simulation was carried out on the basic pancreatic trypsin inhibitor (BPTI) protein for 9ps in a vacuum [160]. From the beginning, the potential for a fruitful relationship between MD and NMR was recognized, and in fact this early simulation (subsequently extended to 96ps) qualitatively agreed with contemporary experimental and theoretical arguments regarding the fluidity of the hydrophobic core [161; 162]. Excitingly, the longest molecular dynamics simulation conducted to date, which sampled 1 millisecond of continuous dynamics, also was performed on the BPTI protein [42]. A wide range of NMR observables can be directly calculated from a molecular dynamics trajectory and compared to experiment, covering dynamical features on a variety of timescales: site-specific measures of flexibility on the ps-ns and μs-ms timescales for both the backbone and sidechains, dihedral angle populations, sidechain rotamers, and interatomic distances are all accessible. Both an advantage and a limitation of NMR measurements is that they are necessarily made on an ensemble of molecules in solution, and understanding the structural origins of ensemble-averaged values is often a challenge. It is this gap in understanding that MD simulations offer the ability to bridge [163].

Historically, the one of the primary NMR observables used for comparison to MD simulations has been the generalized order parameter, S^2, typically used to describe the amplitude of motion of the protein backbone amide bond vector [164; 165]. S^2 is one of several parameters that emerges from the interpretation of NMR spin-relaxation data within the context of the Lipari-Szabo model-free formalism [166; 131]. A complete derivation is outside the scope of the present work, but in brief, the S^2 value reflects the long-time plateau value of the internal correlation function of the motion of a backbone amide vector, under the assumption that internal dynamics and global tumbling occur on distinct, separable timescales. This motion is typically interpreted as "diffusion in a cone", within which the range of S^2 values can be thought of as 0 corresponding to a flexible site and 1 corresponding to a rigid site.

The backbone amide order parameter has a number of deficiencies as a metric for benchmarking simulations. Its relatively small dynamic range tends to result in MD-NMR correlation coefficients being dominated by a small number of highly flexible sites. It does not report on the dynamics of the sidechains, and while equivalent parameters can be calculated for, e.g., the arginine and tryptophan N^ϵ groups, and S^2_{axis} values can be calculated for methyl groups, there are relatively few experimental datasets reporting on these values compared to those reporting on the backbone. Furthermore, the experimental values are highly dependent on assumptions made about the chemical environment of a given nucleus, and on the validity of the timescale-separation assumption, making quantitative comparisons to simulation difficult [167]. More recent MD validation approaches have omitted the model-free formalism and calculated spin-relaxation data directly from long-timescale simulations; however, this is somewhat impractical for large proteins [168]. Residual dipolar couplings (RDCs) have been used as more information-rich sources of experimental data; however, these tend to report on behavior at timescales in the multi-μs regime, which is often out of reach for simulations of large systems on commodity hardware, and the experiments themselves are complex to carry out and high-quality datasets exist for only a few small proteins [169].

The NMR chemical shift value is perhaps the most intuitively appealing as a benchmark for simulations, simply because it is the most easily accessible and least "processed" experimental observable. The physical process that defines a "chemical shift" is relatively straightforward. Briefly: a hypothetical isolated nucleus in a static magnetic field B_0 precesses at the Larmor frequency, which is entirely determined by the characteristic gyromagnetic ratio of the nucleus type:

$$\omega_0 = \gamma B_0 \tag{1.7}$$

However, nuclei are surrounded by electron clouds that circulate in response to the magnetic field, causing the nucleus to experience a slightly lower apparent magnetic field, a process known as nuclear shielding. Moreover, electron motion depends on the details of the surrounding chemical environment, so individual nuclei in a macromolecule may experience different degrees of shielding, which can be expressed as:

$$\omega = \left[1 - \frac{1}{3}\left(\sigma_{xx} + \sigma_{yy} + \sigma_{zz}\right)\right]\gamma B_0 \tag{1.8}$$

where the σ values reflect the diagonal components of the shielding tensor. In solution-state NMR, isotropic tumbling results in the averaging of these distinct directional contributions into a single isotropic value (or more typically, an axially symmetric tensor) rendering them difficult to measure; however, careful experimental and theoretical work suggests that the asymmetry between these components—known as chemical shift anisotropy (CSA)—is distinct for different hydrogen-bonding environments and thus for different secondary structural elements [170; 171; 172], though some older work did not find evidence of this distinction [135]. The calculation of backbone amide order parameters from NMR measurements is often complicated by the need for assumptions about the value of the CSA for ^{15}N nuclei (e.g. [72; 171]); chemical shift values themselves are not constrained by this difficulty, although improvements in solution-state CSA measurements remain of interest in characterizing the local chemical and electronic environment of individual sites in a protein [172].

For practical reasons, chemical shift values are generally not reported as absolute resonance frequencies, but instead as relative frequencies referenced to a standard compound and expressed in parts per million (ppm):

$$\delta = \left(\omega_{obs} - \omega_{ref}\right)/\omega_{ref} \times 10^6 \tag{1.9}$$

which greatly facilitates the comparison of chemical shifts measured at different static magnetic fields.

Furthermore, protein chemical shifts are generally expressed as the sum of two terms, δ_{RC} and δ_{SC}, referring to the random-coil and secondary-shift contributions, respectively. The former is a property of a particular amino acid type and is typically conceptualized as the expected chemical shift for a residue in a generic, completely unstructured environment [173]; the latter is a property of the specific local environment of a particular residue in a folded protein and can be used as a reliable predictor of secondary structure [174]. Major contributors to observed chemical shift values for a given nucleus include hydrogen bonding, backbone dihedral conformation, sidechain rotamers, and ring-current effects due to the circulating π electrons from neighboring aromatic groups [175]. Thus the chemical shift reports with high sensitivity on the local chemical environment of a given nucleus, yet is relatively free of complicating assumptions required for the analysis of the experimental data. Recent advances in the prediction of chemical shifts from static crystal structures have been based on improved quantity and quality of experimental data, improved machine-learning methods, and better insight into the contributions of various types of chemical environment to the shift value, particularly improvements in the modeling of ring-current effects [176; 177; 178; 179; 180; 175; 181]. Chemical shift predictions have recently emerged as valuable tools for the experimental validation of MD simulations [182; 183; 184; 185].

The combination of increasing timescales accessible to simulation and improved capacity to compare simulated and experimental data has resulted in a number of important large-scale benchmarks in the recent past. NMR data have become a standard experimental benchmark for both evaluating force fields [186; 187] and developing novel force field optimizations designed to empirically improve agreement between simulation and experiment [188; 149; 150].

1.3.3 MD simulations as tools for exploring thermal adaptation

The relationship between protein dynamics and thermostability has been extensively explored by a variety of experimental techniques. Unsurprisingly, molecular dynamics simulations have played an important parallel role in providing dynamical interpretations for these experimental observations. A number of simulation studies have been conducted in which the dynamics of homologous proteins from organisms of differing growth temperature were compared to one another as a means of understanding the structural origins of experimentally determined flexibility. Selected examples are briefly reviewed here.

Adenylate kinase In the well-characterized case of the adenylate kinase family, simulations examined the dynamics of salt bridges identified in the crystal structure of the thermophilic but not the mesophilic homolog and studied their persistence over time. Salt bridges found to be persistent in simulation were assumed to contribute to the thermostability of this homolog, a hypothesis that was tested by substituting the salt bridges into corresponding sites of the mesophilic homolog. As predicted, the persistent salt bridges conferred additional thermostability to the mesophilic homolog, whereas the dynamic salt bridge did not [189]. Recent coarse-grained simulation studies comparing mesophilic and thermophilic adenylate kinases observed conformational populations in the mesophile at ambient temperature that more closely matched the thermophilic ensemble at elevated rather than ambient temperature, suggesting at least approximate population of corresponding kinetically competent states [190].

Dihydrofolate reductase The dihydrofolate reductase (DHFR) family has been at the center of the significant recent controversy surrounding the relationship between protein dynamics and catalytic activity following the design of a mutation specifically intended to perturb a dynamic mode thought to contribute to catalysis [47]. Because its chemical mechanism relies on hydrogen tunneling, simulations of DHFR are frequently performed using a quantum mechanics/molecular mechanics (QM/MM) approach to elucidate the possibility of direct coupling between enzyme motion and the chemical reaction coordinate. Simulations have been reported of a “coupled network” of motions over a range of timescales purportedly promoting the enzymatic reaction in the mesophilic homolog [191]. Subsequent work has criticized not only this conclusion but the conceptual edifice of coupling between catalytic activity, dynamics, and thermostability on the basis of simulations in which dynamics are shown not to relate to catalysis and in fact are increased in the thermophile [66]. More recent simulations on the mesophile [49], on a hyperthermophilic homolog [192], and on comparisons between a wild-type and conformationally restricted mutant of a mesophilic homolog [193] have identified putative coupling between protein dynamics and catalysis, although for subtly different definitions of the nature of the effect (a recurring theme in the literature on this topic).

Rubredoxin A number of simulation studies have been conducted on the rubredoxin family. The extent to which the expected flexibility difference between the thermophile and mesophile is observed depends on the quality of the simulation, with early works noting the expected pattern of rigidity in the thermophile [194], while subsequent work found minimal difference between the thermophilic and mesophilic homologs at ambient temperature, although resistance to simulated thermal unfolding was observed in the thermophile [195]. Finally, a third study reported increased flexibility in the thermophile [196]. It is possible that rubredoxin represents a special case, since experimental studies suggest that its primary mechanism of thermostabilization is via an extremely slow unfolding rate, making its stability a kinetic trap rather than an example of classical thermodynamically driven stabilization [64]. (Interestingly, this mechanism has also been observed in RNase HII [197].) Recent studies using a graph-theoretical approach to the prediction of protein dynamics support the notion of increased rigidity in the thermophile [198]; unfortunately, however, this extensively simulated system has not yet been studied with modern molecular dynamics methods, which might aid in resolving the ambiguity that persists from both the experimental and simulation work on this protein family.

Ribonuclease H Finally, several prior MD studies have been conducted on the RNase H protein family. Three of those studies were conducted in the 1990s and report exclusively on the dynamics of ecRNH in comparison to the NMR spin-relaxation data available at the time [132; 199; 200]. Reasonably good agreement is observed between calculated and experimental S^2 order parameters even for trajectories that are very short by modern standards (under 1ns) [132; 199], and for comparison to the NMR-determined structural ensemble [200; 201]; however, incomplete sampling is unsurprisingly observed for a number of residues of interest. An additional two studies from the same time period performed free-energy calculations on thermostabilizing mutants of ecRNH and found surprisingly good agreement with the experimental free energy changes associated with these mutations [202; 203]. A more recent work directly compares the simulated dynamics of ecRNH and ttRNH in the framework of studying the relationship between thermostability and flexibility, but does not attempt to validate the (still fairly short) simulations by comparison to experimental data [204]. In the context of this prior simulation work, the need is emphasized for a comprehensive simulation study conducted in light of recent developments in simulation technology, experimental developments, and structural coverage of the RNase H family.

1.4 Dissertation overview

This dissertation presents a series of studies conducted in the pursuit of an improved understanding of the complex relationships between protein dynamics, activity, and thermostability in the ribonuclease H protein family. The organizing principle of the work presented herein has been the close coupling between molecular dynamics simulations and nuclear magnetic resonance spectroscopy observations, permitting both validation of the

MD trajectories by rigorous comparison to experiment, and improved interpretation of the dynamics observed by NMR in light of the incomparable spatial and temporal resolution afforded by MD. The molecular origins of the increased thermostability and decreased catalytic activity in the thermophilic homolog have remained obscure despite decades of effort, and the work presented here aims to shed light on the subject from the unique perspective gained by long-timescale simulations.

The first study in this dissertation focuses on the dynamics of a widely conserved structural feature of the RNase H family known as the handle loop, which is dispensable for activity under some conditions, but nevertheless known to be involved in substrate binding and an important locus of thermostability contributed by specific residue substitutions in the thermophilic ttRNH. Comparative analysis of molecular dynamics simulations of a total of five homologous RNase H families from organisms of varying preferred growth temperature reveals systematic differences in handle-region conformational dynamics. Previous NMR observations of handle-region dynamics in ecRNH and ttRNH are integrated into an interpretive framework derived from simulations of this larger family. These results illustrate the utility of combined MD-NMR studies in elucidating the effects of particular amino acid residues on molecular adaptation to features of the bulk environment.

The second study takes as its inspiration the discovery of a set of three mutations that collectively confer increased activity on ttRNH at minimal cost to thermostability. The dynamic consequences of these activating mutations are rationalized in the context of observed differences in sidechain orientation in the wild-type ecRNH and ttRNH simulations. Importantly, these effects are distinct from effects on handle-loop conformation.

The third study focuses on the dynamics of the ecRNH active site. First, the dynamics of carboxyl- and carbonyl-containing sidechains in molecular dynamics simulations of ecRNH were compared to those inferred from recent NMR experiments quantifying motion of these residues at the ps-ns and μs-ms timescales. These residues are of particular interest because the conserved RNase H active site contains four carboxylate groups. These results suggest that the active site residues are rigid in the ps-ns timescale while undergoing substantial conformational exchange upon Mg^{2+} binding. This may be interpreted as evidence in favor of electrostatic preorganization for binding the first metal ion, coupled to dynamic reorganization at longer timescales. This work illustrates the advantage of combined MD-NMR studies for understanding the dynamic prerequisites for enzymatic catalysis.

The fourth study investigates the utility of extremely long-timescale molecular dynamics simulations using the recently developed special-purpose hardware platform Anton for further exploring the multi-μs dynamics experienced by RNase H proteins. Unexpectedly, local unfolding in the handle region of several RNase H homologs at long timescale is observed, highlighting the need for careful validation of force fields intended for use in this type of high-performance simulation. However, the ctRNH homolog did not experience unfolding and conformational preferences in sidechains known to be important for substrate binding could be analyzed; this simulation represents the longest known molecular dynamics trajectory for an enzyme from a thermophilic organism.

Finally, the fifth study presented in this dissertation reports on the examination of the ctRNH protein by NMR. This RNase H homolog has not previously been characterized by NMR and is predicted based on simulation to experience conformational dynamics distinct from those shared by ecRNH and ttRNH, possibly representing an alternative evolutionary mode of thermal adaptation.

Chapter 2

Materials and methods

2.1 Molecular dynamics simulations

2.1.1 System preparation

All systems for simulation were prepared using a common protocol to ensure comparability among the entire dataset. For each initial protein structure, protonation states for titratable residues were assigned either by experimental measurement (for ecRNH [205]) or by prediction using the H++ [206] pKa predictor. Unless otherwise specified, all simulations were performed at a pH of 5.5 to recapitulate the conditions used in prior NMR experiments on ecRNH and ttRNH [72; 73]. Crystallographic water molecules were removed from all structures prior to solvation using the Maestro tool, version 8.5 or 9.1. All systems were solvated with the TIP3P water model [207] and proteins were described with the AMBER99SB force field [208] unless otherwise specified.

PDB structures used for initiating trajectories and as platforms for mutagenesis are listed, along with their resolutions and any system-specific preparation steps, in Table 2.1.

It should be noted that in some cases the crystal structures differ from the wild-type proteins, and these differences were not reverted by modeling prior to simulation. In particular, in 1RIL the C-terminal tail of ttRNH is not resolved; this region has been shown to be highly dynamic [72] and examination of dynamically averaged chemical shifts in this region did not reveal significant deviations from experiment [184], suggesting that this region is dispensable for modeling the behavior of the structured portion of the protein. Additionally, NMR experiments were performed on the cysteine-free version of the protein, but simulations were carried out on the cysteine-containing 1RIL structure without modification. In 3H08, the crystal structure corresponds to the cysteine-free mutant, which was simulated without reversion of these mutations. For all of the retroviral proteins and 2QK9, the N-terminus of the crystal structure is not the natural N-terminus of the protein; in all cases the residues present in the structure were simulated.

Table 2.1: **Structurally characterized RNase H homologs.**

PDB ID	Protein	Source organism	Res (Å)	Preparation and comments
2RN2 [95]	ecRNH	*Escherichia coli*	1.48	—
1RNH [94]	ecRNH	*Escherichia coli*	2.00	Agrees less well with NMR order parameters compared to 2RN2-initiated trajectories
1RDD [120]	ecRNH	*Escherichia coli*	2.80	*E. coli* WT with single bound Mg^{2+} ion
1GOA [136]	ecRNH iG80b	*Escherichia coli*	1.90	*E. coli* glycine insertion mutant
1GOC [136]	ecRNH iG80b A77G	*Escherichia coli*	2.00	*E. coli* glycine insertion/A77G double mutant
1RIL [69]	ttRNH	*Thermus thermophilus*	2.80	—
2E4L [102]	soRNH	*Shewanella oneidensis*	2.00	—
3H08 [101]	ctRNH	*Chlorobium tepidum*	1.60	Missing residues in handle and active-site loop modeled in from 1RIL (chosen due to the presence of the glycine insertion in both proteins)
2QK9 [119]	hsRNH	*Homo sapiens*	2.55	Substrate was removed and catalytically inactivating D210N mutation reversed using Maestro 8.5
3K2P [209]	hivRNH	HIV	2.04	Inhibitor and bound metal ions removed in Maestro 9.1; chosen as the HIV structure with lowest rmsd to the unbound state (PDB ID 1HRH) with the active-site loop resolved
3P1G [210]	xmrvRNH ΔC	XMRV	1.60	Helix C and handle region deletion mutant of XMRV RNase H domain with single bound Mg^{2+} ion
3V1O [211]	xmrvRNH	XMRV	1.88	Full length XMRV RNase H domain with no bound ion
4E89 [212]	xmrvRNH	XMRV	2.60	Full length XMRV RNase H domain with single bound Mg^{2+} ion
2LSN [213]	pfvRNH	PFV	N/A (NMR)	Full length PFV RNase H domain (containing helix C and the handle)

2.1.2 Computational mutagenesis

Computational mutagenesis on the structures in Table 2.1 was performed in Maestro version 9.1 for solvent-exposed sites or MODELLER v9.5 for packed sites. For MODELLER models, residues with at least one heavy atom within a 5Åsphere of the site of interest were considered mobile, while the distal portions of the structure were held fixed to minimize the perturbation introduced by the mutation; if this procedure failed, the entire structure was allowed to be mobile.

For ttRNH dG80, no crystal structure was available, so a model was produced in MODELLER using 1RIL and 2RN2 as templates. Point mutations were then made using this model as a starting structure.

2.1.3 Simulations on commodity hardware

Simulations were performed using Desmond Academic release 3 or source release 2.4.2.1 [214]. Unless stated otherwise, proteins were described with the Amber99SB force field [208], solvated with TIP3P water [207] in a cubic box with a 10 Å buffer region from solute to box boundary, and neutralized with Cl^- ions. Bonds to hydrogen atoms were constrained using the M-SHAKE algorithm [215]. Simulations containing Mg^{2+} ions used the Aqvist parameter set [216]. Electrostatics were calculated with the PME method using a 9Åcutoff; results were not affected by the use of more conservative electrostatics parameters. All simulations used a 2.5fs inner timestep on a 1-1-3 RESPA cycle and were performed in the NVT ensemble using a Nosé-Hoover thermostat after equilibration to constant box volume for 5ns in the NPT ensemble. All simulations described in this work were run for 100ns unless otherwise noted.

2.1.4 Simulations on special-purpose hardware (Anton)

Simulations were initiated from the last frames of either 100ns or 1μs (for ecRNH and ttRNH) trajectories previously calculated using Desmond source release 2.4.2.1 or 3.4.0.1. All simulations were carried out using Anton software version 2.11.0. Except where otherwise stated, simulations used a 2.0fs inner timestep on a 1-1-3 RESPA schedule using the Multigrator integrator with a Nosé-Hoover thermostat. In no case does choice of integrator affect the results.

2.1.5 Trajectory analysis

Handle-region dynamics were monitored using a reaction coordinate consisting of the Cartesian distance between the residues equivalent to W85 and A93 in ecRNH; values greater than 10Å were considered to reflect an open state. Order parameters were calculated using the equation [217]:

$$S^2 = \frac{1}{2}\left(3\sum_{i=1}^{3}\sum_{j=1}^{3}\langle\mu_i\mu_j\rangle^2 - 1\right) \tag{2.1}$$

in which μ_i and μ_j represent the x, y, and z components of a unit vector $\vec{\mu}$ in the direction of a given chemical bond. This represents the long-time limit of the angular reorientational correlation function for a given bond vector.

Standard trajectory analysis—extraction of dihedral angles, hydrogen bond occupancy, inter-residue contacts, secondary structure, etc.—was performed within the VMD environment using custom Python extensions.

2.2 Sequence analysis

Sequences of bacterial RNase H domains were collected from InterPro entry IPR002156 [218] (in May 2012) and annotated for source organism growth temperature using the Integrated Microbial Genomes database [219]. Sequences that were redundant or did not contain a handle loop were removed and the remaining sequences aligned to the four available bacterial structures using PROMALS3D [220].

Evolutionary coupling analysis was performed using the direct-coupling method as implemented in the EVfold webserver [221].

2.3 Chemical shift predictions

Chemical shift predictions were performed as described [184]. For shift calculations described in the present work, predictions were performed using the SPARTA+ [175] program. For every 10th frame of 100ns and 1μs trajectories of ecRNH and ttRNH, the coordinates of the protein were extracted and minimized into the AMBER03 force field [222] by 200 steps of steepest-descent optimization using the simulation toolkit almost-1.0.4 [223], a combination which has been shown to produce robust prediction quality for a range of chemical shift prediction tools applied to PDB structures [180]. This approach was intended to avoid noise in predictions due to suboptimal crystal-structure geometry, but produced insignificant effects on the final predictions from MD-generated structures. To facilitate comparison between predicted and experimental shifts, all experimental values were rereferenced using SHIFTCOR [224].

2.4 Nuclear magnetic resonance spectroscopy

2.4.1 Sample preparation

Samples of [U-^{13}C,U-^{15}N] *C. tepidum* RNase H were prepared using standard protein overexpression and purification methods. A pAED4 plasmid containing the coding sequence of the cysteine-free ctRNH mutant was a generous gift from the laboratory of Prof. Susan Marqusee (UC Berkeley). This ctRNH gene was subcloned into the pET-47b+ vector (Genscript, Inc.), which provides an N-terminal His-tag sequence and a kanamycin-resistant selectable marker, for consistency with laboratory protocols for purification of ecRNH and its point mutants. A cleavage site recognizable by TEV protease was introduced for the purpose of removing the His tag after purification. (This results in an additional glycine residue at the N-terminus of the protein, which is a common outcome of His-tagged purification and is not expected to affect the dynamics of the protein.) A point mutation N88R was introduced (Genscript, Inc.) and this protein purified in parallel to the wild-type ctRNH.

E. coli BL21(DE3) cells were transformed with pET-47b+ vector and used to inoculate 3mL cultures in rich media plus kanamycin grown overnight. Approximately 2mL of these cultures were used to inoculate 1L cultures in M9 minimal media, supplemented with 10mg biotin, which was prepared with 1g/L $^{15}NH_4Cl$ and 4g/L $U^{13}C$-glucose (Cambridge Isotope Laboratories). These cultures were grown to an OD of approximately 0.6 (7 hours growth) before induction of protein expression. Cells were harvested after approximately 3.5 hours of expression.

Cells were lysed by sonication and purified using standard methods for preparation of His-tagged protein. Crude lysate was centrifuged to remove insoluble cellular debris and then applied to a freshly charged nickel-containing HisTrap HP column (GE Biosciences) and eluted using a gradient of 5-500mM imidazole concentration. Protein-containing fractions as detected by absorbance at 280nm and verified by SDS-PAGE were pooled and dialyzed into a buffer containing 50mM Tris-Hcl and 0.5mM EDTA, pH 7.5, and exposed to 1mL of commercial TEV protease in the presence of 1mM DTT for at least 12 hours at room temperature. TEV and remaining impurities were then removed by application to tandem Q-HP and Heparin-HP HiTrap columns (GE Biosciences); the Q column was removed and the heparin column eluted with a gradient of 50-1000mM NaCl. Finally, remaining uncleaved protein was removed by pooling protein-containing fractions and applying them again to the His column. Both WT and N88R ctRNH samples bind nonspecifically to the His column and were eluted at approximately 100mM imidazole. All column purification steps were performed at room temperature using a Bio-Rad chromatography system. Identity and purity of the resulting protein was confirmed by mass spectrometry.

Final yields for the two proteins were 26mg/L for WT ctRNH and 18mg/L for N88R ctRNH. To prepare NMR samples, half of each sample was buffer-exchanged into 100mM sodium *d3*-acetate, 10% D_2O, pH 5.5, and concentrated to approximately 500μL total volume. Final concentrations for samples used in NMR assignment experiments were 0.95mM WT ctRNH and 0.73mM N88R ctRNH.

2.4.2 Backbone triple-resonance assignments

A Bruker DRX600 NMR spectrometer equipped with a CryoProbe was used to acquire all NMR spectra. Data were collected at 299K and internally referenced using 1mM DSS. Standard backbone triple-resonance assignment experiments, minimally consisting of ^{15}N-HSQC, HNCACB, and HN(CO)CACB, were collected for both proteins. In the case of the wild-type ctRNH, HNCO and HACONH spectra were collected as well. Future work on this system may necessitate additional data collection, particularly if comparison to predictions motivates interest in the H_α assignments of the N88R mutant.

2.4.3 Processing and data analysis

All spectra were processed using the NMRPipe software suite [225] and analyzed using the visualization and assignment tools Sparky [226] and CCPNmr Analysis [227]. Additional visualization was performed using the Python module NMRglue [228].

Chapter 3

Dynamics of the RNase H handle loop

3.1 Introduction

Key features of the structure of RNase H are illustrated in Figure 3.1; of particular note is the region of the protein encompassing helices B and C and the following loop, which is known as the handle region or the basic protrusion due to its density of positively charged residues. Although some RNase H homologs lack helix C and the handle loop altogether [92], and ecRNH has been shown to retain some activity when this region is deleted [229], biochemical evidence clearly associates the region with substrate binding [230; 231; 232]. A naturally handle-less homologous subdomain from the HIV retroviral reverse transcriptase lacks activity in isolation, but an insertion mutant containing the ecRNH handle sequence regains activity under some conditions [230; 231]. Alanine scanning mutations in helix C and the handle loop identify several conserved tryptophan residues critical for binding and reveal that neutralizing positively charged residues in the handle additively disrupts binding affinity [232]. Moreover, crystal structures of the *Homo sapiens* homolog (hsRNH) in complex with substrate show extensive contacts between the DNA strand of the substrate and residues located in helix C and the handle region [119]. Additionally, NMR relaxation measurements suggest that the handle region and a second long loop near the active site have similar rates of motion on the $\mu s - ms$ timescale, suggesting a coupled motional process [73].

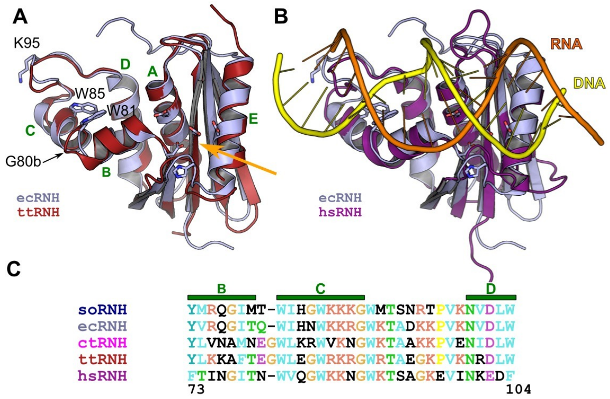
A
K95
D
A
W85
W81
E
C
G80b
B
ecRNH
ttRNH
B
RNA
DNA
ecRNH
hsRNH
C
B
C
D
soRNH YMRQGIMT-WIHGWKKKGWMTSNRTPVKNVDLW
ecRNH YVRQGITQ-WIHNWKKRGWKTADKKPVKNVDLW
ctRNH YLVNAMNEGWLKRWVKNGWKTAAKKPVENIDLW
ttRNH YLKKAFTEGWLEGWRKRGWRTAEGKPVKNRDLW
hsRNH FTINGITN-WVQGWKKNGWKTSAGKEVINKEDF
73
104

Figure 3.1: **Key features of RNase H structure and sequence.**(A) Structural superposition of ecRNH (light blue; PDB ID 2RN2) and ttRNH (red; PDB ID 1RIL). Helices are labeled with green letters and key residues in the handle region and active site (orange arrow) are shown as sticks. (B) Superposition of the ecRNH structure (light blue) with the substrate-bound complex of the hsRNH protein (purple; PDB ID 2QK9), illustrating the position of the handle region interacting with the DNA strand (yellow) of the DNA:RNA hybrid substrate. (C) Sequence alignment of helices B, C, and the handle loop for all five homologs studied.

Two sites near the handle region have been previously identified as major contributors to the differences between ecRNH and ttRNH. First, an inserted glycine, numbered G80b, is present in ttRNH in the junction between helices B and C. NMR studies of ecRNH and ttRNH show increased chemical exchange in the handle region for ttRNH, indicating motion on a $\mu s - ms$ timescale [72]. Reciprocal mutations reveal that the glycine insertion mutant ecRNH iG80b possesses thermophile-like relaxation behavior and significantly impaired catalytic activity; on the other hand, the deletion mutant ttRNH dG80b possesses mesophile-like relaxation behavior, although its activity does not increase [73;

69]. Second, a site at the tip of the handle loop with a conserved left-handed helical conformation in Ramachandran space is occupied by a lysine in ecRNH and a glycine in ttRNH. The ecRNH K95G mutant increases thermostability by 1.9 kcal/mol, likely due to the elimination of the steric strain associated with non-glycine residues in left-handed conformations [98].

Despite this extensive history, the relationships between dynamics, thermostability, and enzymatic activity in the RNase H family remain obscure. Herein we integrate previous NMR observations of handle-region dynamics in ecRNH and ttRNH into an interpretive framework derived from molecular dynamics simulations of all handle-region-containing family members of known structure. These results illustrate the utility of combined MD-NMR studies in elucidating the effects of particular amino acid residues on molecular adaptation to features of the bulk environment.

3.2 Two-state behavior in the handle region

3.2.1 Two-state behavior in wild-type RNases H

We begin with the three proteins containing an arginine or lysine residue at position 88 at the end of helix C: soRNH, ecRNH, and ttRNH. The motion of the handle region in each protein is monitored by a reaction coordinate consisting of a simple Cartesian distance metric between the $C\alpha$ atoms of A93 at the tip of the handle loop and W85 as an anchor point on helix C (ecRNH residues and numbering), as illustrated in Figure 3.2B and Figure 3.4 and plotted as a function of simulation time for representative trajectories in Figure 3.3. These three proteins share a conserved dynamic mode in which two distinct handle conformations are observed, an open and closed state. The open state is populated by soRNH and ecRNH at lower temperatures, while elevated temperatures simply equalize the populations of each state, as expected. In contrast, ttRNH predominantly occupies the closed conformation at all temperatures studied. Notably, the corresponding distances in the crystal structures of all three proteins lie between the two conformations observed in the simulations (Figure 3.2C), possibly due to the presence of crystal contacts in that region (Figure 3.5A) or an inability to model both states during crystallographic refinement. The thermostability of the ecRNH protein has been extensively studied by mutagenesis; a survey of these ecRNH mutant structures, though dominated by contact-stabilized intermediate conformations, also identifies examples of both the open and closed conformations (Figure 3.5B). Preference for the open conformation among the two more active homologs suggests that this may be the conformation competent for substrate binding. We hypothesize that ttRNH is reliant on thermal fluctuations to access the open conformation on a timescale exceeding that studied here. This pattern is reminiscent of observations previously made in triose phosphate isomerase[41], dihydrofolate reductase[233], and adenylate kinase [234], in which simulations suggest rapid, nanosecond-timescale sampling of partially activated conformations, but a stable fully activated conformation is suggested by experiment to be accessible only at millisecond timescales.

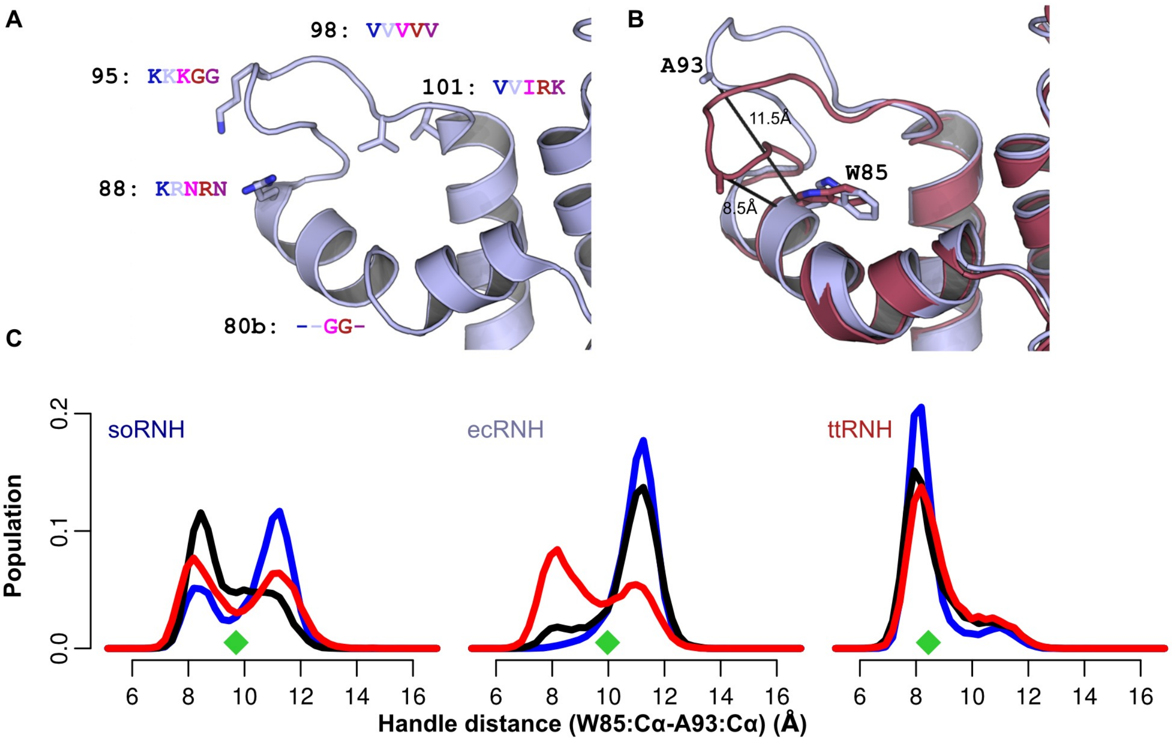

Figure 3.2: **Dynamics of the RNase H handle region as a function of temperature.** (A) Key residues modulating handle region dynamics and their identities in each homolog (soRNH, dark blue; ecRNH, light blue; ctRNH, magenta; ttRNH, red; hsRNH, purple). (B) Representative conformations from the ecRNH trajectory of the open (blue) and closed (brown) states, illustrating the Cartesian distance metric used as a reaction coordinate. (C) Temperature dependence of soRNH (left), ecRNH (middle), and ttRNH (right) handle-region dynamics illustrating the relative populations of the closed and open states at 273K (blue), 300K (black), and 340K (red). Measurements of the distance metric from each crystal structure are shown as green diamonds.

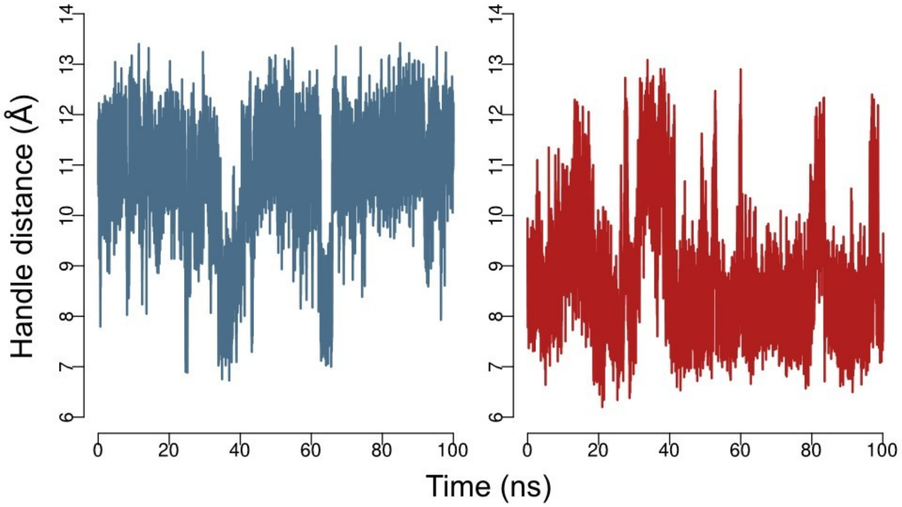

Figure 3.3: **Timecourses of handle region dynamics for ecRNH and ttRNH.**The fluctuations of the handle-region distance metric as a function of time are shown for ecRNH (left; blue) and ttRNH (right; red) for the 300K trajectories, representing 100ns of simulation time.

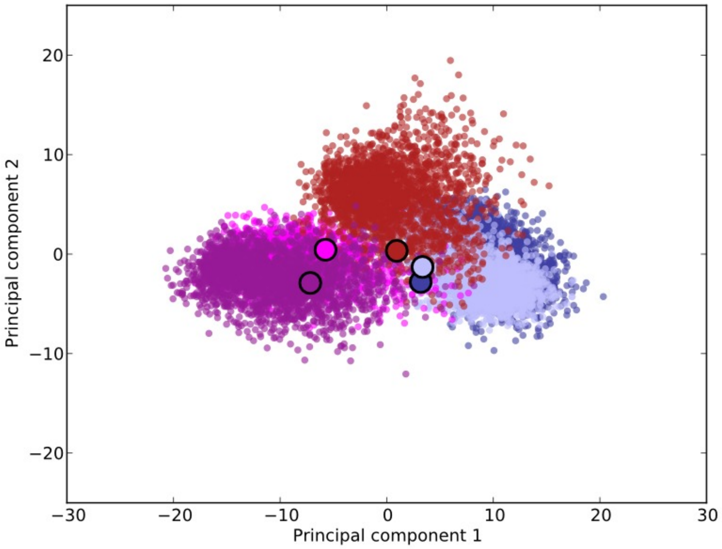

Figure 3.4: **Principal components analysis of the handle loop for all five RNase H proteins.** PCA analysis on the C_α Cartesian coordinates of the handle loop, corresponding to residues G89 to N100 in ecRNH, was carried out on the 300K trajectories of all five wild-type proteins. Projections onto the first two principal components are shown for soRNH (dark blue), ecRNH (light blue), ctRNH (magenta), ttRNH (red), and hsRNH (purple); crystal structures are indicated as filled circles. The first principal component axis describes the difference between single-state and two-state proteins, while the second describes the difference between the open and closed states. Collectively these two principal components account for 89% of the variance in the dataset.

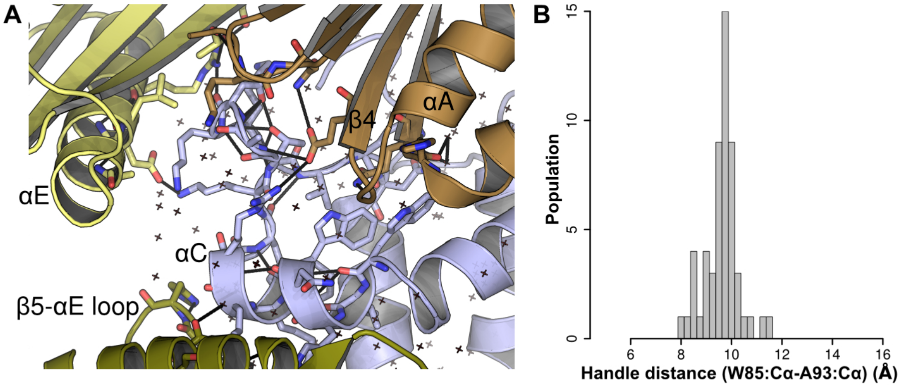

Figure 3.5: **Crystal contacts identified in ecRNH structures.** (A) The crystal-packing environment surrounding the ecRNH handle region (blue) in 2RN2. Symmetry mates are shown in green, yellow, and brown; the local hydrogen bonding network is shown as black lines. (B) Distribution of handle-distance measurements in 54 chains representing 32 PDB structures of the ecRNH protein.

3.2.2 Comparison of experimental and simulated NMR observables for the handle loop

The ecRNH and ttRNH simulations can be validated by comparison to experimental NMR data. Calculated S^2 order parameters, reflecting amplitude of local motion, are in good agreement with the experimental values for both proteins (Figure 3.6). In addition, we have previously shown that simulation-derived chemical shift predictions reflecting dynamic conformational averaging perform significantly better than predictions from the static crystal structures in reproducing experimental chemical shift data for ecRNH and ttRNH [184]. This agreement is particularly significant because chemical shifts, especially those of protons, are highly sensitive to ring-current effects from the orientation of aromatic groups, which are plentiful near the handle loop. The accuracy of dynamically averaged predictions of chemical shifts for these two proteins (Figure 3.7) supports the hypothesis that the motions observed in the 300K simulations recapitulate motions observed experimentally. The handle loop typically shows below-average RMSDs to the experimental chemical shift values (Table 3.1), suggesting that this particularly dynamic region is reasonably well-sampled.

Table 3.1: **Chemical shift RMSDs for flexible regions of ecRNH and ttRNH**

	ecRNH				ttRNH			
Protein region	C_α	HN	H_α	N	C_α	HN	H_α	N
Handle (MD)	0.45	0.42	0.31	2.34	0.38	0.42	0.25	1.97
Handle (Xray)	0.38	0.47	0.27	2.51	1.16	0.51	0.28	4.02
Average (MD)	0.70	0.36	0.25	2.25	0.68	0.35	0.23	2.17
Average (Xray)	0.74	0.39	0.25	2.51	1.00	0.44	0.32	2.70

RMSDs for the C_α and N Sparta+ chemical shift predictions to experimental values for ecRNH[235] and ttRNH[72]. The improvement due to dynamic averaging is particularly good in the handle region for the relatively low-resolution ttRNH structure, while the ecRNH values are within the magnitude of the error of the predictor.

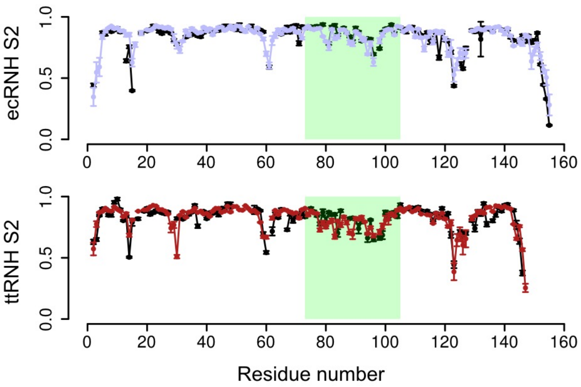

Figure 3.6: **Predicted vs. experimental backbone amide order parameters.** (A) Comparison between experimental [135] (black) and predicted (blue) S^2 order parameters for ecRNH. Helices B, C, and the handle region are highlighted in green. (B) Comparison between experimental [72] (black) and predicted (red) order parameters for ttRNH. Correlations as determined by Pearson's R are 0.89 and 0.74 respectively; the lower correlation for ttRNH is likely due to the fact that the experimental values were acquired at 310K using a cysteine-free form of the protein to avoid undesirable thiol chemistry. Experimental values have been rescaled by the slope of a linear regression to the simulated values for visualization.

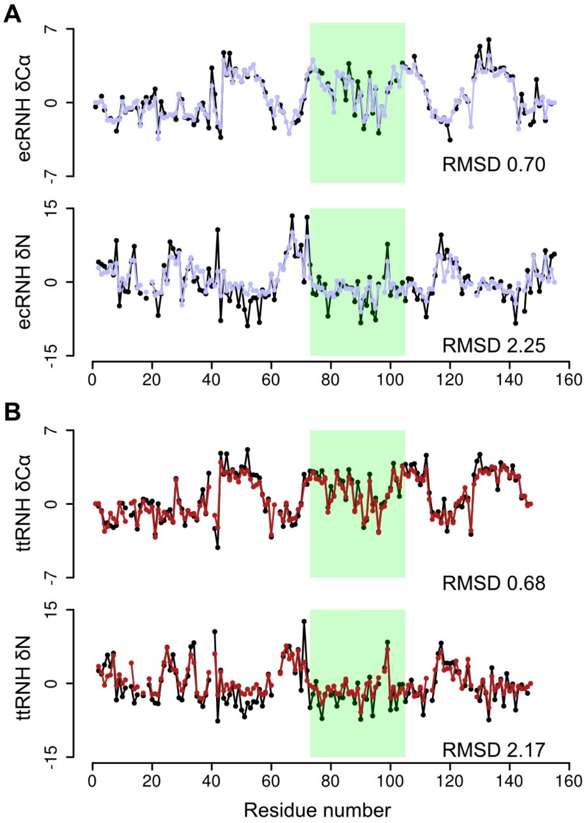

Figure 3.7: **Dynamically averaged chemical shift predictions.**(A) Comparison between experimental (black) and predicted (blue) secondary chemical shift values for the nuclei with the smallest (C_α) and largest (N) RMSD values among those predicted in [184] for ecRNH. Helices B, C, and the handle region are highlighted in green. (B) Comparison between experimental (black) and predicted (red) secondary chemical shift values for ttRNH. Predicted values are reproduced from [184]. Values are plotted as secondary chemical shifts (deviation from random-coil value for each residue); RMSDs are calculated using the absolute shift values.

3.2.3 Reinterpretation of NMR measurements based on simulations

Previous NMR relaxation measurements on ecRNH and ttRNH produced estimates of the relative free energies of major and minor conformational states, summarized in the free-energy diagram in Figure 3.8A [73]. This landscape was constructed based on the observations that a) the ecRNH and ttRNH crystal structures closely resemble one another, and b) those structures do not appear to be in a binding-competent conformation. However, this result was perplexing because the putatively binding-competent state was more highly populated in the less-active ttRNH. Population estimates from the simulations suggest an alternative interpretation (Figure 3.8): that the minor state of ecRNH at 300K is equivalent to the major state of ttRNH, and vice versa; thus, a mirrored version of our original free energy profile is likely a better representation of the experimental data. While it is unlikely that such short simulations reproduce equilibrium behavior, the overall picture of a conserved dynamic process with a larger activation barrier in ttRNH is consistent with previous observations [72].

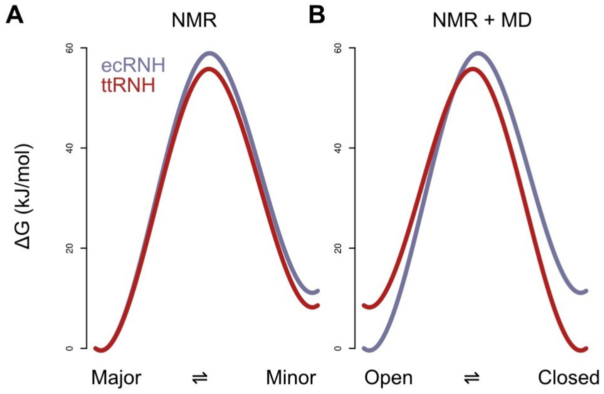

Figure 3.8: **Free-energy landscapes for putative major and minor states for ecRNH and ttRNH.** (A) The free-energy diagram constructed under the assumption that both ecRNH and ttRNH share a major state that is incompetent for substrate binding [73]. (B) A revised diagram, inspired by the simulated populations, in which the landscape for ttRNH is mirrored, suggesting that the (closed) minor state of ecRNH is equivalent to the major state of ttRNH, and the (open) major state of ecRNH is equivalent to the minor state of ttRNH.

3.3 An alternative mode of substrate binding

3.3.1 Single-state behavior in wild-type RNases H

Two proteins in our data set, ctRNH and hsRNH, contain an asparagine at residue 88, where the other proteins contain arginine or lysine. The natively Asn-containing proteins do not exhibit two-state behavior, but instead show a single peak for the handle-region metric, centered around the crystal structure value and broadening with increasing temperature (Figure 3.9A). Although hsRNH was crystallized in complex with substrate and might be thought to occupy a distinct handle-region conformation due to substrate interactions, the average all-to-all $C\alpha$ RMSD between the handle regions in the 300K trajectories of hsRNH and ctRNH, which was crystallized without substrate, is only 1.04 Å.

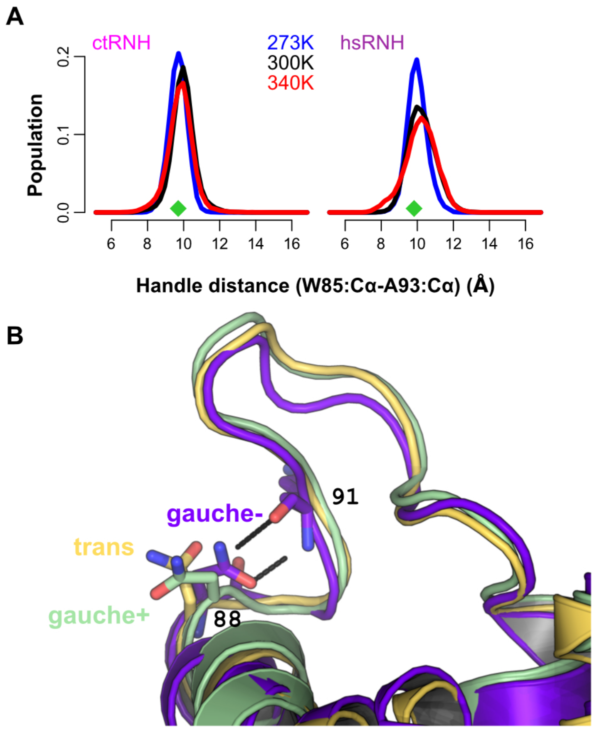

Figure 3.9: **Dynamics of the handle region in the presence of Asn at position 88.** (A) Temperature-dependent populations for the two natively Asn-containing proteins—ctRNH (left) and hsRNH (right)—illustrating the presence of a single conformationaln distribution centered around the crystal-structure position (green diamonds) whose basin broadens with increasing temperature. (B) The presence of Asn at position 88 permits the formation of two highly stable hydrogen bonds to the backbone of residue 91 when Asn occupies the *gauche-* χ_1 rotamer; the ecRNH R88N mutant, whose trajectory samples all three states, is shown here.

3.3.2 Determination of handle region dynamics by a single residue

To explore the effects of asparagine and arginine on handle-region behavior, we made mutations at this site for all five proteins. In four cases, the resulting mutants are stable under the simulation conditions at 300K for 100ns, but hsRNH N88R requires two additional stabilizing mutations: in the prokaryotic proteins, a pair of well-conserved residues, Y73 and W104, anchor the interface between helices B and D; in hsRNH, both are replaced by Phe. The absence of the additional hydrogen bonding contributions in hsRNH N88R disrupts the interfaces between helices B, D, and A (Figure 3.15A-B); however, the triple mutant hsRNH F73Y/N88R/F104W is stable and shows dynamics similar to those observed for the prokaryotic homologs.

The dynamic consequences of substitutions at position 88 are clearly shown in Figure 3.10: when Arg or Lys occupies this site, the handle region shows two-state behavior, while Asn produces a single handle-distance peak centered roughly between the open and closed states for the two-state systems. In both wild-type and mutant proteins containing Asn, a dominant *gauche-* χ_1 rotameric state for this residue is observed in which the sidechain amide forms two hydrogen bonds to the backbone carbonyl and amide of the neighboring residue at position 91 (Figure 3.9B). By contrast, Arg 88 is highly flexible and forms only transient, often water-mediated hydrogen bonds with its neighbors; the sidechain order parameter for Arg 88 in ecRNH has been measured as around 0.2 at 300K, and this low value is well-reproduced in simulation [236].

The effects on backbone flexibility of the reciprocal mutations at position 88 are relatively complex. Difference in calculcated S^2 between Arg or Lys-containing proteins and Asn-containing proteins are shown in Figure 3.11. Natively Arg or Lys containing proteins (top row) consistently have more rigid helix B with their native residue than with Asn. However, ttRNH has a significantly more rigid helix C with Asn. In all five proteins, the

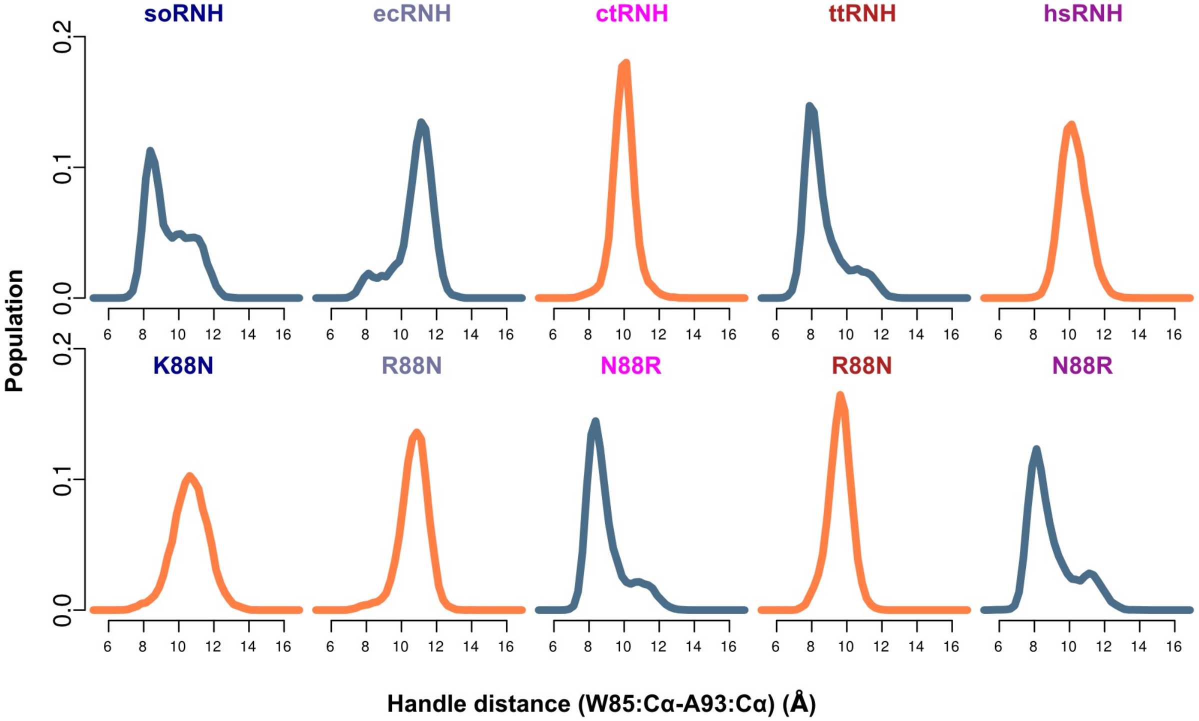

Figure 3.10: **Handle-region behavior for reciprocal mutants at position 88.** The top row shows wild-type proteins and the bottom shows each protein's corresponding mutant. Distributions shown in blue indicate the presence of a positively charged residue at position 88 (Arg in all cases except soRNH, which natively contains Lys) and distributions shown in orange indicate the presence of Asn at this position. All simulations were carried out at 300K.

glycine-rich loop is more rigid with the native position 88 residue than with the reciprocal mutant, regardless of the identity of the residue. This suggests that these two loops, both of which form interactions with substrate, may have coupled dynamics which only become detectable in simulation when relatively large perturbations are introduced by mutation.

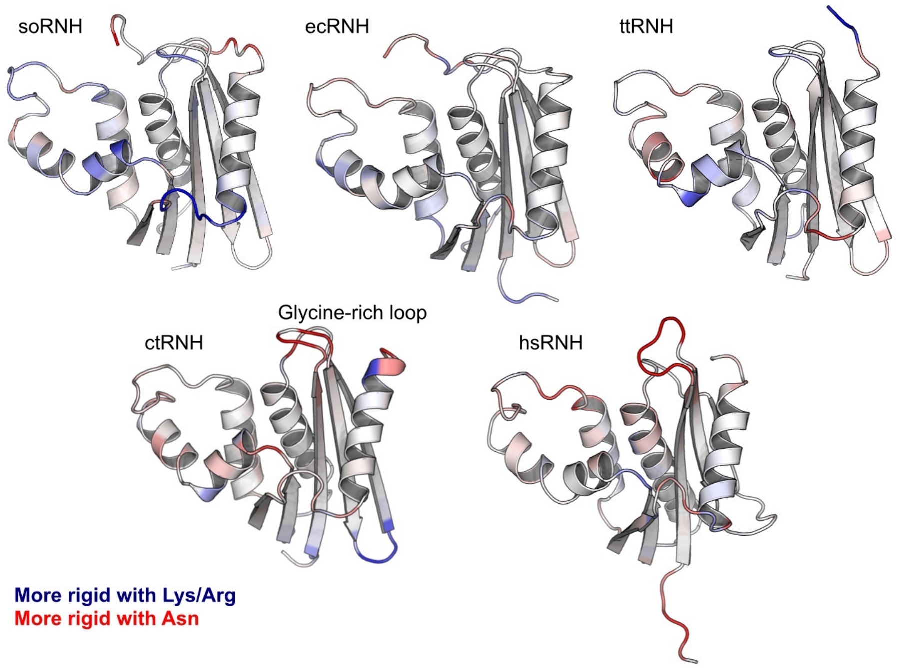

Figure 3.11: **Effects of mutations at position 88 on backbone flexibility.** Proteins in the top row natively contain Arg or Lys; proteins in the bottom row natively contain Asn. Each protein is colored by the difference in calculated S^2 between its Arg- or Lys-containing trajectory and its corresponding Asn trajectory; thus, regions shown in red indicate greater rigidity with Asn and regions shown in blue indicate greater rigidity with Arg or Lys. These representations report on the same 100ns 300K trajectories shown in Figure 3.10.

3.4 Tuning handle region populations

3.4.1 Mutations in known substrate-binding residues

Initial efforts at manipulating the relative populations of the open and closed states focused on residues already known to be involved in substrate binding. In particular, residue W81, one of two strongly conserved tryptophan residues in helix C, is known to form direct contact with substrate in the hsRNH complex (Figure 3.1) through stacking of its indole ring. The W81A mutation in ecRNH is known to reduce enzymatic activity (P. Robustelli, unpublished) and this residue is observed in prior simulation work to exhibit differential rotamer dynamics between ecRNH and ttRNH [74]. Simulation of the W81A ecRNH mutant yielded rapid convergence to a stable closed state coupled to the burial of the alanine sidechain (Figure 3.12).

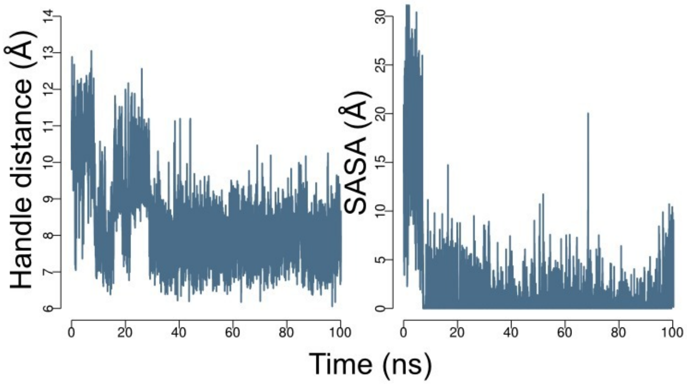

Figure 3.12: **Handle-region dynamics in ecRNH W81A mutant.** The behavior of the handle-distance metric as a function of simulation time is shown at left; the solvent-accessible surface area of the A81 sidechain is shown at right.

However, because the effects of changes in dynamics are difficult to decouple from the effects of the loss of specific substrate interactions, additional mutations were sought at sites not directly in contact with substrate.

3.4.2 Mutations in distal residues

In seeking additional sites to target for *in silico* and experimental mutagenesis, chemical shift predictions were used to identify residues exhibiting a distinctive change in shift distribution in the open and closed states. Although the conformational change associated with handle region motion is quite subtle, it was hoped that this approach would result in candidate mutants with a relatively straightforward experimental readout to validate predicted changes in preferred handle conformation.

The remaining four residue positions highlighted in Figure 3.2A are identified by the simulations as critically important in determining the relative populations of the open and closed states. Three of these sites—the glycine insertion (G80b), Val 98, and Val 101 (ecRNH residues and numbering)—form the borders of a hydrophobic spine linking helices C and D through the two conserved Trp residues involved in direct substrate contacts. In ecRNH, rotamer jumps at the two valine sites correlate with both predicted chemical shift and the handle-distance metric (Figure 3.13 and 3.14). The remaining site, Lys 95, resides at the tip of the handle loop and requires a left-handed helical backbone conformation. Strategic substitutions at these sites allow us to rationally manipulate the relative populations of the open and closed states in both native and mutant proteins with a positively charged residue at position 88.

Position 98 is highly conserved as a Val among prokaryotic RNases H that possess handle regions, underscoring its functional significance, despite the lack of direct contact between its sidechain and substrate. The mutant ecRNH V98A abrogates the observed rotamer transitions and populates a predominantly closed conformation (Figure 3.14A).

In ecRNH, rotameric transitions of Val 101 induce subtle changes in local packing throughout the hydrophobic spine, potentially stabilizing the open conformation. To produce a ttRNH mutant with increased population of the open state, we therefore made

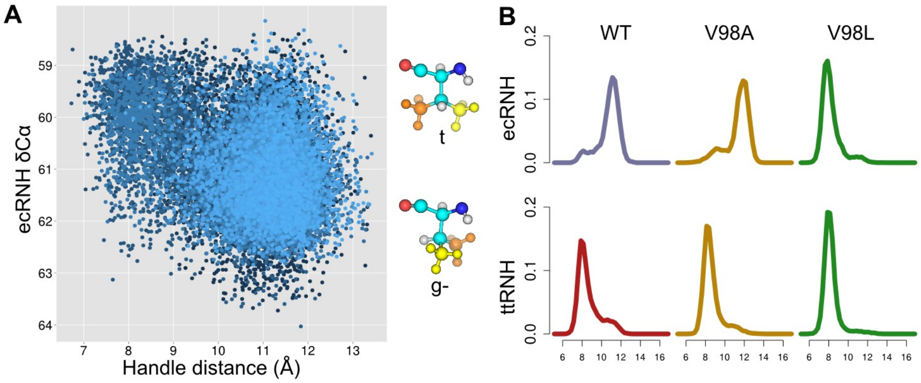

Figure 3.13: **Coupling of handle-region dynamics to V98.** (A) Correlation between the handle distance metric and the C_α chemical shift of V98 in ecRNH. Structures at right indicate the most common V98 rotamer giving rise to the corresponding chemical shift value. Points are colored from dark to light blue to reflect the timecourse of the trajectory. (B) Effects of V98 mutants on ecRNH (top) and ttRNH (bottom).

reciprocal mutations at this position in both the presence and absence of the inserted Gly at position 81 and the left-handed Gly residue at position 95, which is occupied by a Lys in ecRNH. The results of these mutations are summarized in Figure 3.14B. In brief, the mutations work in concert; while no single mutant significantly increases open state population, a ttRNH dG80/G95K/R101V triple mutant populates the open state at a level of about 40%, compared to about 10-15% for the wild-type and dG80 enzymes. Conversely, an ecRNH K95G/V101R/Q105E mutant enriches population of the closed state relative to wild type. (In this case a double mutant was necessary to provide the Arg with an equivalent to its native hydrogen-bonding partner.) The success of these mutations in altering the local conformational equilibrium underscores the importance of this hydrophobic cluster.

Notably, corresponding mutations in the context of the hsRNH F73Y/N88R/F104W triple mutant produce the same effects on its open-closed dynamics. The wild-type hsRNH protein lacks a glycine insertion but contains a Gly at position 95 and a Lys at position 101, similar to the ttRNH protein. The quintuple mutant obtained by the additional G95K/K101V substitutions significantly increases the population of the open state relative

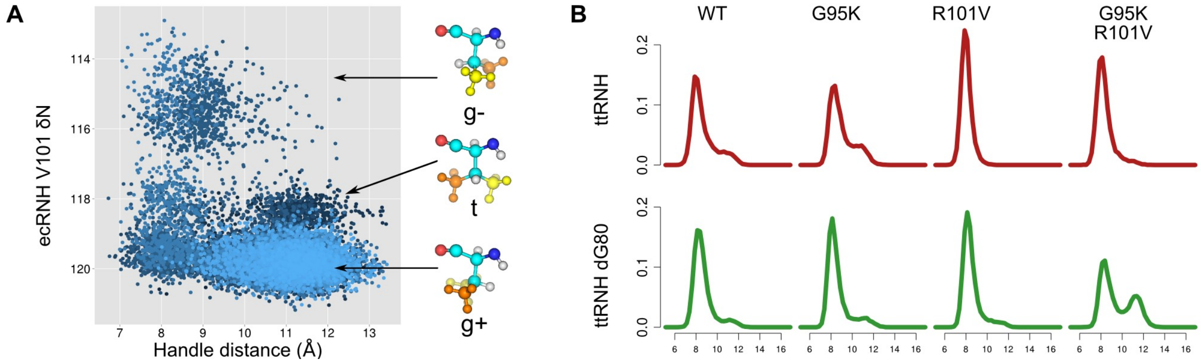

Figure 3.14: **Coupling of handle-region dynamics to V101.** (A) Correlation between the handle distance metric and the N chemical shift of V101 in ecRNH. Structures at right indicate the most common V101 rotamer giving rise to the corresponding chemical shift value. Points are colored from dark to light blue to reflect the timecourse of the trajectory. (B) Coupled effects of mutations at positions 95 and 101 in ttRNH (top) and ttRNH dG80 (bottom). Only ttRNH dG80 G95K/R101V shows a population of the open state significantly enriched compared to wild-type ttRNH.

to the triple mutant. Similarly, ctRNH dG80/N88R is predicted from its sequence—K95, I101—to predominantly populate the open state in solution. This protein, like hsRNH, required reengineering of the interface between helices B and D to form a stable structure (Figure 3.15C-D); the modified form of the protein behaves as predicted, populating the open conformation more frequently with the native K95/I101 residues than with the mutant G95/R101 (Figure 3.16).

3.5 Sequence analysis

The sites identified by structural and dynamic considerations to play an important role in handle-region motions also exhibit weak trends among available RNase H sequences in favor of dual modes of thermal adaptation. Notably, sites previously identified as relevant to thermostabilization among RNase H proteins—positions 80b and 95—play an important role in cooperatively determining relative populations of open and closed states. For po-

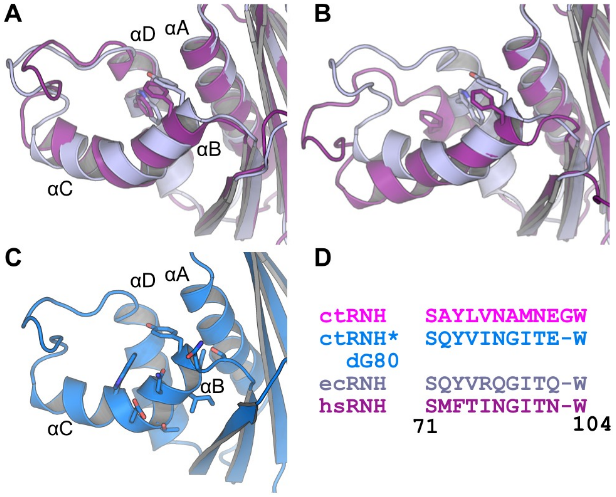

Figure 3.15: **Mutations introduced to stabilize the helix B-helix D interface in N88R mutants** (A) Superposition of ecRNH (blue) and hsRNH (purple), illustrating the phenylalanines mutated in hsRNH* to their homologous bacterial residues. (B) Destabilized conformation of hsRNH N88R in the absence of the hsRNH* mutations F73Y/F104W. (C) Model of ctRNH dG80 mutant ctRNH*, illustrating the mutations made to the packing interface in helix B to construct a stable context into which to make the N88R mutation. (D) Sequence of ctRNH* dG80 helix B.

sitions 80b, 95, and 101, the residues that contribute to increased closed-state population are favored among sequences annotated as derived from thermophilic organisms (Figures 3.17, 3.18), suggesting that adaptation to high-temperature environments directly trades off against population of the open state. These results suggest that mesophilic organisms tolerate thermally destabilizing non-glycine residues in the left-handed dihedral conformation structurally required at position 95 due to their effects on relative open-state population.

In addition, among bacterial proteins containing handle loops, the frequency of ocurrence of Asn at position 88 is higher among those sequences annotated as having a thermophilic source organism than among those annotated as being derived from mesophiles (Figure 3.18). A suppressor screen for thermostabilizing mutations of soRNH, which natively contains Lys at this position, identified K90N as thermostabilizing by 0.7 kcal/mol with only a 9% decrease in activity relative to the wild-type protein [97], consistent with

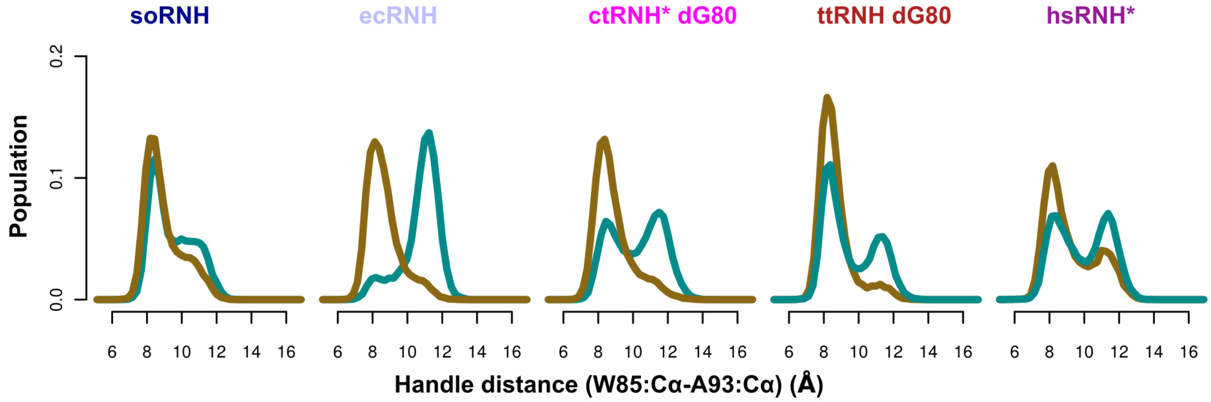

Figure 3.16: **Manipulation of relative populations by coupled mutations at positions 95 and 101.** For all arginine- or lysine-containing proteins other than soRNH, mutants containing G95 and R101 (brown) populate the closed state more frequently than those containing K95 and V101 (cyan), regardless of the wild-type residues at these positions. The natively N88-containing proteins, ctRNH and hsRNH, both required additional mutations to stabilize the interface between helices B and D, as detailed in Figure 3.15.

our observations by computational mutagenesis that these reciprocal mutations are mostly nondisruptive and are easily accommodated in the local environment.

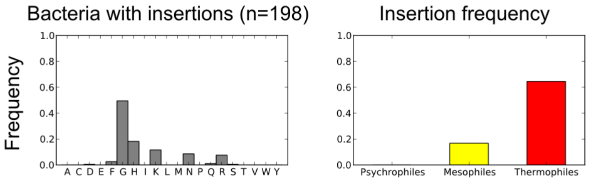

Figure 3.17: **Residue frequencies in the glycine-insertion position.** Distribution of residues among the 198 bacterial RNase H domain sequences identified as possessing an insertion (left); frequency of insertion as a function of growth temperature annotation of the source organism (right).

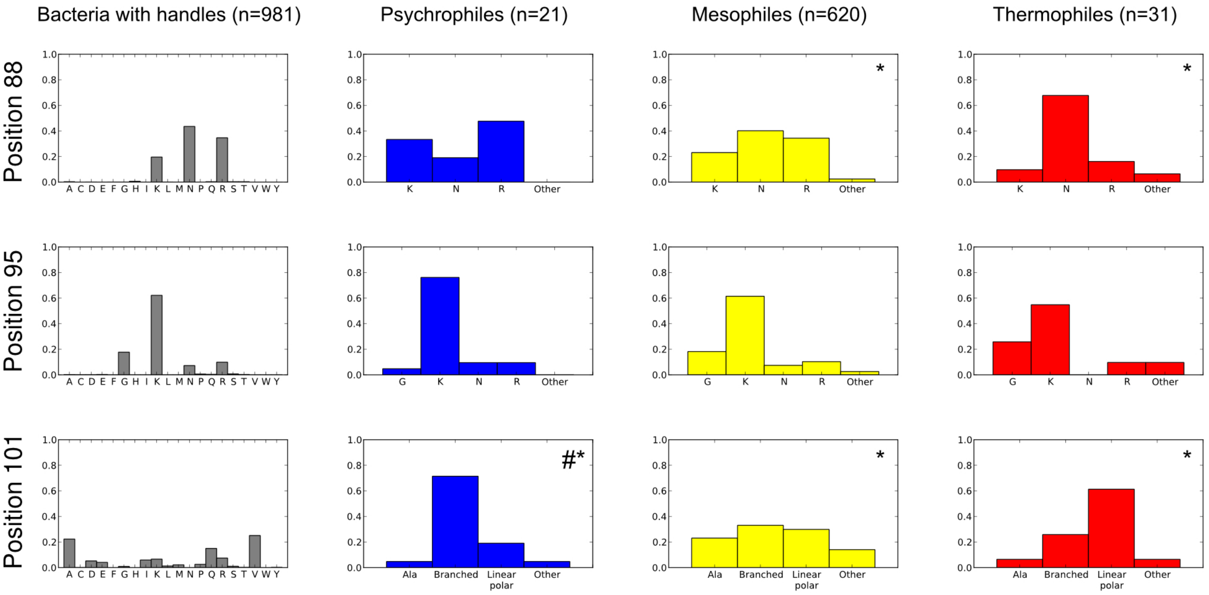

Figure 3.18: **Residue frequencies in sites identified as significant determinants of handle-region dynamics.** Distribution of residues at each of positions 88, 95, and 101 among bacterial RNase H sequences from all organisms, and as a function of growth temperature annotation.For position 101, residues have been clustered into four categories: alanines, branched amino acids (isoleucine, leucine, valine), linear and polar amino acids (arginine, lysine, glutamate, glutamine), and other amino acids. For positions 88 and 101, the notation * indicates a distribution significantly different from uniform, and the notation # indicates a distribution significantly different from the overall dataset (χ^2 test with Bonferroni-corrected significance level of p<0.003). Mean percent sequence identities for each category are 54% (overall), 62% (psychrophiles), 55% (mesophiles), 51% (thermophiles).

Chapter 4

Sidechain dynamics and activating mutations of ttRNH

4.1 Introduction

Three dynamic loops outline the substrate-interaction surface of RNases H. In addition to the handle loop already discussed, two additional loops contribute to the formation of specific substrate interactions: the glycine-rich loop and the β5-αE loop, also known as the active-site loop. The behavior of specific substrate-interacting sidechains in or near these loops has previously been examined by Nikola Trbovic's prior simulation work on RNase H [74], in which three distinct dynamics reflecting changes in sidechain orientation are identified that differ between ttRNH and the more active ecRNH.

Mutations in ttRNH have previously been identified by genetic screening that result in increased catalytic activity at ambient temperature [237]. The locations of the mutations—A12S, K75M, and A77P—are indicated in Figure 4.1A. The effects of these mutations are summarized in Table 4.1.

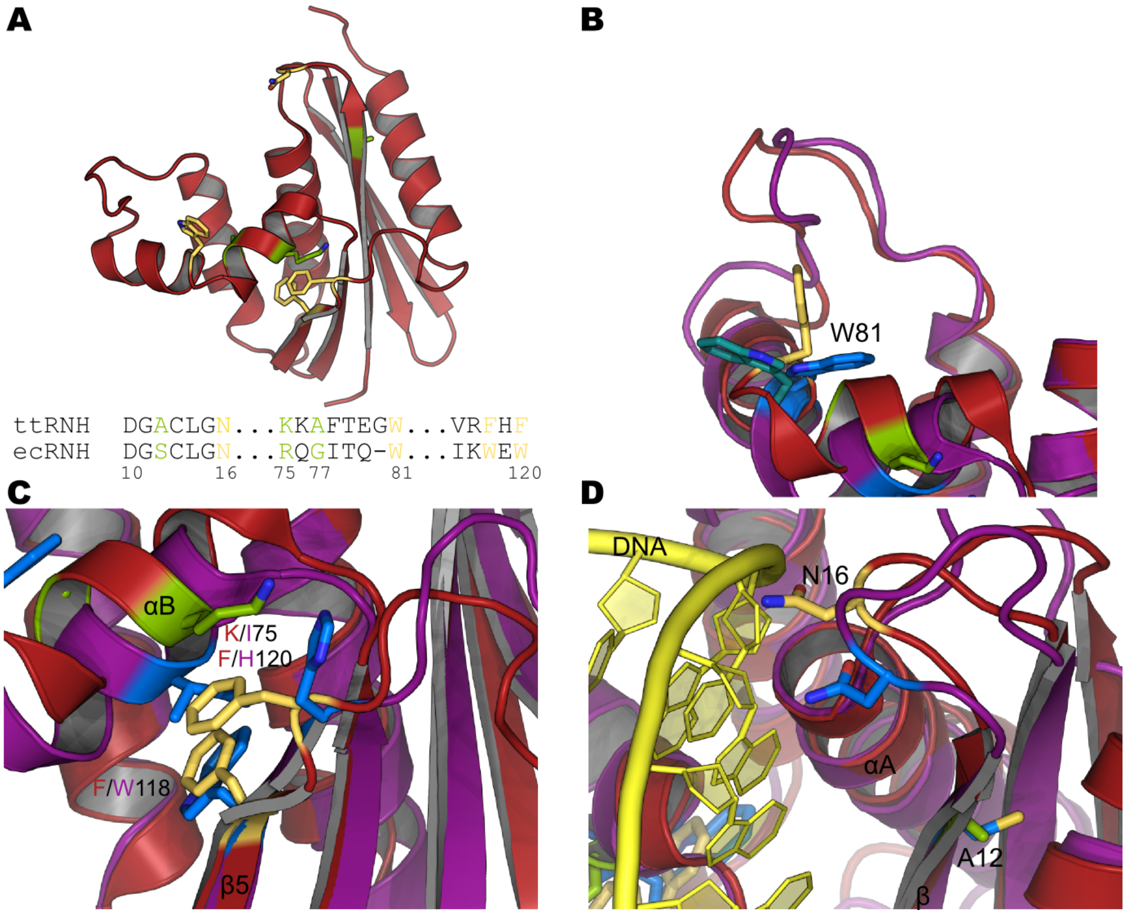
A
B
ttRNH DGACLGN...KKAFTEGW...VRFHF
ecRNH DGSCLGN...RQGITQ-W...IKWEW
10 16 75 77 81 120
W81
C
D
αB
K/I75
F/H120
F/W118
β5
DNA
N16
αA
A12
β

Figure 4.1: **Activating mutants and sidechain dynamics identified in ttRNH.** (A) The three residues identified as contributing to increased catalytic activity in ttRNH [237] are shown in green. The residues previously observed [74] to show differential dynamics in ttRNH and ecRNH are shown in yellow. (B) Superposition of ttRNH (red/yellow) and hsRNH (purple/blue), illustrating the conformational space of the W81 rotamer. The ttRNH crystal-structure conformation (yellow) is not populated in simulation; instead, the productive conformation (blue) and an unproductive conformation (cyan) are in equilibrium. (C) Superposition of ttRNH (red/yellow/green) and hsRNH (purple/blue) at the interface between helix B and sheet 5, illustating differences in local packing. Positions are annotated with the ttRNH residue in red text and the hsRNH residue in purple. (D) Superposition of ttRNH and hsRNH in the glycine-rich loop, where the productive conformation (blue) of N16 is represented by hsRNH and an unproductive conformation that may create a steric clash with substrate is represented by ttRNH (yellow).

Table 4.1: **Properties of catalytically activating ttRNH mutants.**

Enzyme	K_m (μM)	k_{cat} (min^{-1})	T_m (° C)
ttRNH WT	7.8	4.9	77.4
A12S	1.5	4.5	76.0
K75M	6.1	8.0	78.2
A77P	3.3	7.3	71.7
A12S/K75M	1.1	12	76.8
A12S/K75M/A77P	1.1	27	75.0

Properties of mutant ttRNH enzymes identified by genetic screening as contributing to increased activity at ambient temperature (30°C). Table reproduced in part from Tables 1 and 2 of [237].

In the present work, 100ns molecular dynamics simulations are described of homology models based on the ttRNH crystal structure (PDB ID 1RIL) for the double mutant A12S/K75M and the triple mutant A12S/K75M/A77P. Because introducing a proline residue into the second turn of a short helix is potentially difficult to model, two independent homology models were generated for the triple mutant and trajectories were initiated from both starting conformations. Herein the dynamic consequences of these activating mutations are rationalized in the context of observed differences in sidechain orientation in the wild-type ecRNH and ttRNH simulations. Importantly, these effects are distinct from effects on handle-loop conformation, as none of the activating-mutant trajectories result in increased population of the handle-loop open state relative to wild-type ttRNH. Although the tradeoff between activity and stability is often analyzed in terms of overall backbone flexibility, two out of the three activating mutations in ttRNH specifically perturb the rotamer dynamics of relatively well-packed hydrophobic groups, inducing a series of compensating changes that do not perturb the overall average flexibility of the backbone.

Table 4.2: **Populations of minor G15 conformation in activating mutants of ttRNH.**

Enzyme	Population
ttRNH WT 1	1.7%
ttRNH WT 2	4.2%
A12S/K75M	13.6%
A12S/K75M/A77P 1	4.4%
A12S/K75M/A77P 2	9.0%

Populations of the minor G15 backbone dihedral conformation, which involves a transition to the left-handed helical region of Ramachandran space, in wild-type and mutant ttRNH. Although the population distributions overlap, the trend is clearly in favor of increased population in the mutants.

4.2 Dynamics of the glycine-rich region

Residues Gly15 and Asn16 in the glycine-rich loop are notably flexible both experimentally and in simulation in ecRNH. By contrast, ttRNH is significantly more rigid at 300K, and becomes flexible only at elevated temperatures. The backbone flexibility reflected in relatively low S^2 order parameters for these residues is associated with occupancy of a minor conformation in Ramachandran space by G15, which results in a change in orientation for the sidechain of N16. The major conformation for this loop, which is represented in almost all crystal structures of bacterial RNases H, results in an orientation in which the N16 sidechain would sterically clash with substrate; however, in the minor G15 backbone conformation, N16 occupies an orientation that closely resembles that found in the hsRNH complex (Figure 4.2). This conformation is also similar to that observed in *Bacillus halodurans* in complex with substrate, despite the fact that the substrate orientation in this complex differs significantly from hsRNH and from an orientation that would be necessary for interactions with the handle [118]. The populations of this state are low at 300K even for the activating mutants, but nevertheless are increased relative to wild-type (Table 4.2) and are correlated with decreased order parameters in the loop (Figure 4.3).

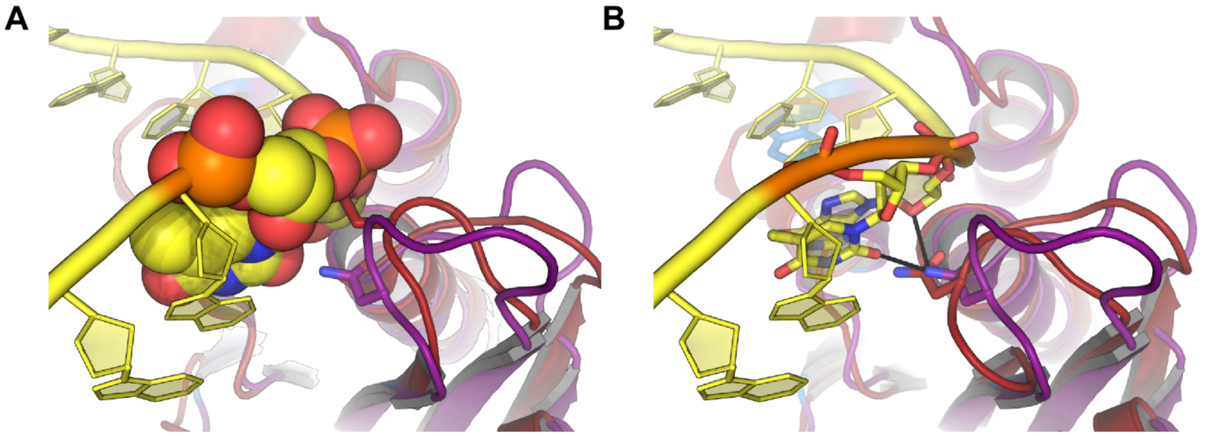

Figure 4.2: **Conformations of the glycine-rich loop.** (A) The ttRNH crystal structure is shown in red; this conformation of N16 is in steric conflict with the DNA strand (yellow) of the substrate in the superposed hsRNH complex. (B) The hsRNH (purple) conformation of N16 forms specific hydrogen-bonding interactions (black lines) with substrate. The minor conformation of the ttRNH (red) backbone orients the N16 sidechain as observed in the bound conformation.

In both ttRNH mutant simulations, the flexibility of this loop is increased relative to wild-type. Because it is unlikely that the K75M mutant would substantially change the dynamics of a loop approximately 20Å away, and the more flexible ecRNH has Ser at position 12, it is reasonable to conclude that the dynamics of this loop are coupled to the A12S mutation. Kinetics measurements suggest that A12S has an effect primarily through K_m, which is consistent with the suggestion that the mutation results in increasing the ability of this loop to assume a binding-competent conformation.

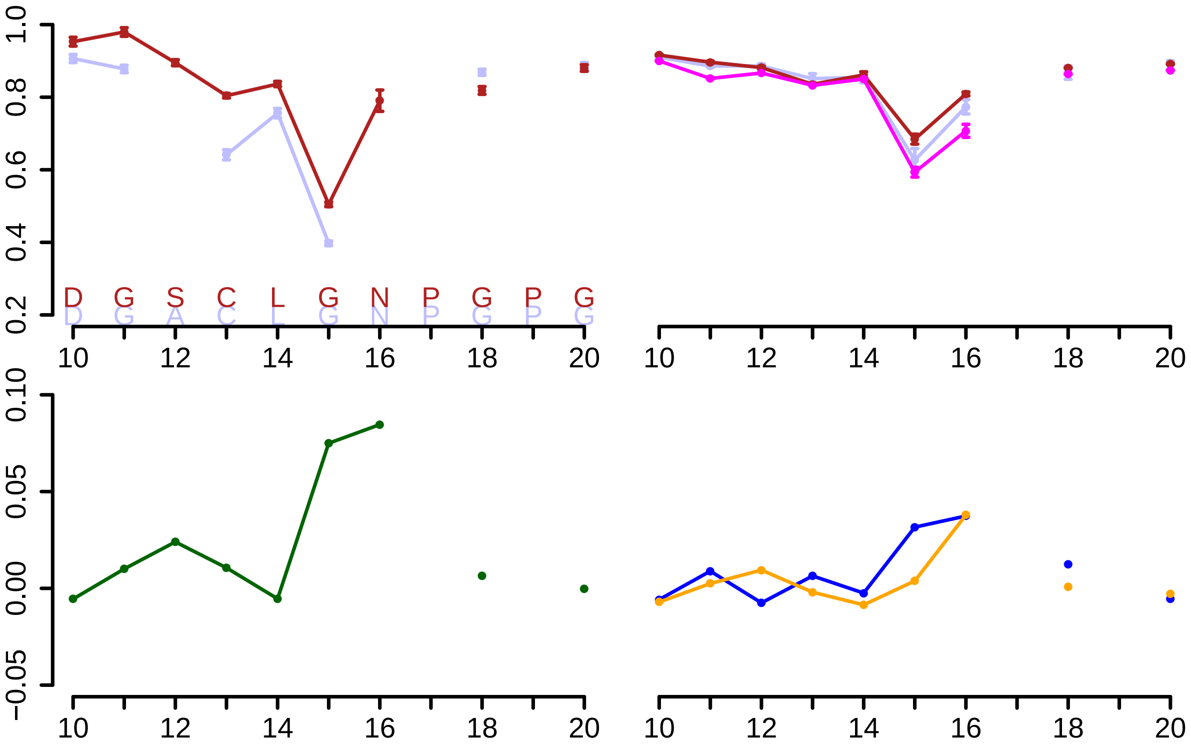

Figure 4.3: **Order parameters for ecRNH, ttRNH, and activating mutants in the glycine-rich loop.** (Top left) Experimental S^2 order parameters for ecRNH (blue) and ttRNH (red). (Top right) Simulated order parameters for ecRNH at 300K (blue), ttRNH at 300K (red), and ttRNH at 340K (magenta). The pattern of increased rigidity in ttRNH is reproduced in simulation, though the simulations systematically overestimate the order parameter value relative to experiment. (Bottom left) Difference in order parameters for A12S/K75M (green) relative to WT ttRNH. (Bottom right) Difference in order parameters for two trajectories of the A12S/K75M/A77P triple mutant (blue, yellow), illustrating increased flexibility in this loop in the presence of the A12S mutation.

4.3 Hydrophobic packing in the helix B-β 5 interface

The packing of the hydrophobic interface between helix B and β-sheet 5 is significantly different in crystal structures of RNases H in the apo state compared to the hsRNH complex. In particular, aromatic groups (Phe in ttRNH, Trp in ecRNH) at positions 118 and 120 are oriented "down", pointing away from the substrate-binding site, in both ecRNH and ttRNH; however, a 180° χ_2 rotation orients the equivalent H120 residue in hsRNH "up", so that its ring nitrogen interacts with the backbone of the RNA strand of the bound substrate. Notably, more distantly related RNases H, including the subdomain found in HIV reverse transcriptase[92] and the atypical RNase H [238] , occupy the "up" conformation even in the absence of substrate.

In wild-type ttRNH, a concerted rotamer transition is observed in which F118 and F120 cooperatively rotate from the "down" to the "up" conformation (Figure 4.4A). Although phenylalanine lacks a ring nitrogen with which to form substrate interactions, and the sidechain is not close enough to substrate for ring-stacking, the resulting conformation is nevertheless intriguing due to its greater similarity to the structure of the complex. While the simulations do not provide sufficient statistics for analyzing the rate of this transition, it is accessible within 100ns in one of three independent trajectories. By contrast, the equivalent residues in ecRNH—W118 and W120—never sample the "up" conformation within 100ns. Instead, ecRNH requires either much longer timescales or elevated pH in order to occupy this conformation, and in no case does W118 reorient alongside W120. In ecRNH the transition is facilitated by opening of the active-site loop immediately adjacent to β-sheet 5, but this does not seem to be strictly required in wild-type ttRNH.

In the A12S/K75M and A12S/K75M/A77P ttRNH trajectories, the transition of F120 to the "up" state occurs much more readily. Notably, the coupling between F118 and F120 is consistently broken; in no K75M-containing trajectory does F118 stably change

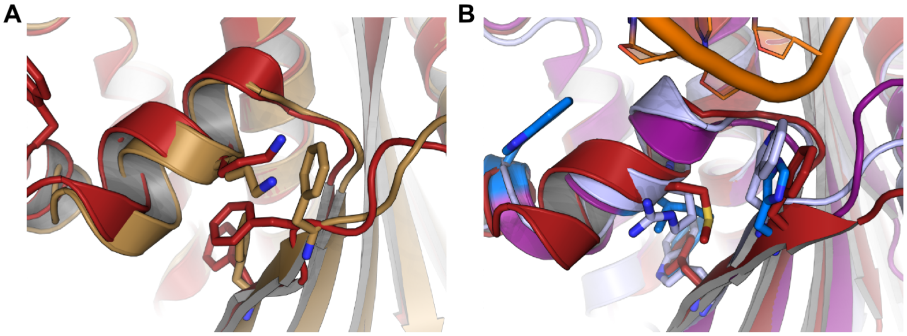

Figure 4.4: **Conformations of the helix B-β 5 interface.** (A) Two conformations accessible to wild-type ttRNH. The crystal structure conformation is shown in red; the alternative conformation after a concerted transition for F118 and F120 is shown in yellow. (B) Conformations represented in the hsRNH complex (purple/blue), in a 1μs trajectory of ecRNH (light blue), and in all ttRNH trajectories featuring a K75M mutation (red). In ecRNH only W120 experiences a rotamer transition, in contrast to wild-type ttRNH; in ttRNH mutants containing K75M, the motions of F120 are similarly decoupled from F118.

orientation. In all cases M75 packs onto the F118 ring in a conformation reminiscent of that observed for the equivalent Ile in hsRNH (Figure 4.4B).

A small but systematic difference is observed between the wild-type ttRNH and its activating mutants in the dynamics of the neighboring active-site loop. The loop is slightly more flexible in the presence of mutations, but particularly in the A12S/K75M double mutant (Figure 4.5).

Characterization of the mutant enzymes suggested that both the double K75M/A77P and the triple A12S/K75M/A77P ttRNH mutants exhibit small differences in near- and far-UV circular dichroism spectra compared to either wild-type ttRNH or the any of the single mutants, implying a subtle difference in the packing of aromatic groups despite little change in secondary structure [237]. It seems likely that facilitating the transition of F120 to the "up" conformation is the mechanism through which the K75M mutation, alone or coupled to A77P, increases catalytic activity in ttRNH. (Because A77P also induces a change in aromatic environment, it is possible that both mutations are necessary in order to produce

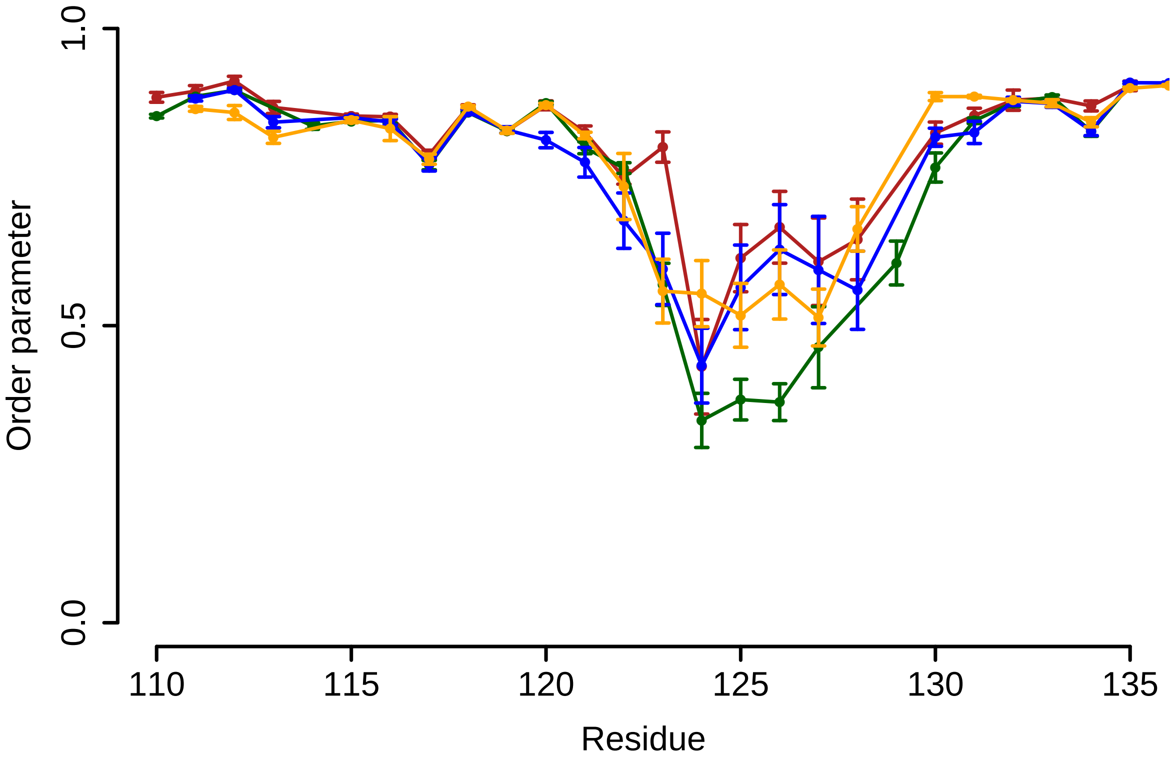

Figure 4.5: **Computed order parameters for wild-type ttRNH and it activating mutants in the active-site loop.** Order parameters are shown for wild-type ttRNH (red), A12S/K75M (green), and A12S/K75M/A77P (blue and yellow).

a difference of observable magnitude, since neither mutation alone shows perturbed CD spectra.) Because the K75M mutation does not have an effect on the K_m value (Table 4.1), it seems as if the F120 rotamer flip is not required for substrate binding; however, it may contribute to maintaining the precise orientation of substrate in the active site required for efficient catalysis.

4.4 Rotamer dynamics of tryptophan 81 in helix B

4.4.1 Rotamer dynamics in wild-type RNases H

Trp 81, located at the N-terminus of helix C, is highly conserved among handle-region-containing bacterial RNases H. Examination of the hsRNH-substrate complex [119] reveals specific interactions through sidechain interactions as well as water-mediated hydrogen bonds through the indole ring nitrogen. Three conformations of W81 are observed among RNase H crystal structures, distinguished by their χ_2 rotamers (Figure 4.6A): the *trans* rotamer, occupied in the hsRNH structure as well as most ecRNH structures; the *gauche-* rotamer, occupied as a minor conformation in ttRNH, and the *gauche+* rotamer, occupied in the ttRNH crystal structure. This last rotamer is likely to be an artifact of crystal contacts in the region, as it is quickly abandoned in all simulations initiated from this conformation and rarely revisited, and clearly clashes with substrate. Interestingly, in a mutant in which the inserted G80b immediately preceding W81 in ttRNH has been added to the ecRNH sequence (denoted ecRNH iG80b), W81 occupies the *gauche-* rotamer in the corresponding crystal structure [136]. By superposition with the hsRNH structure, this rotamer is less well-positioned to accommodate substrate (Figure 4.6B).

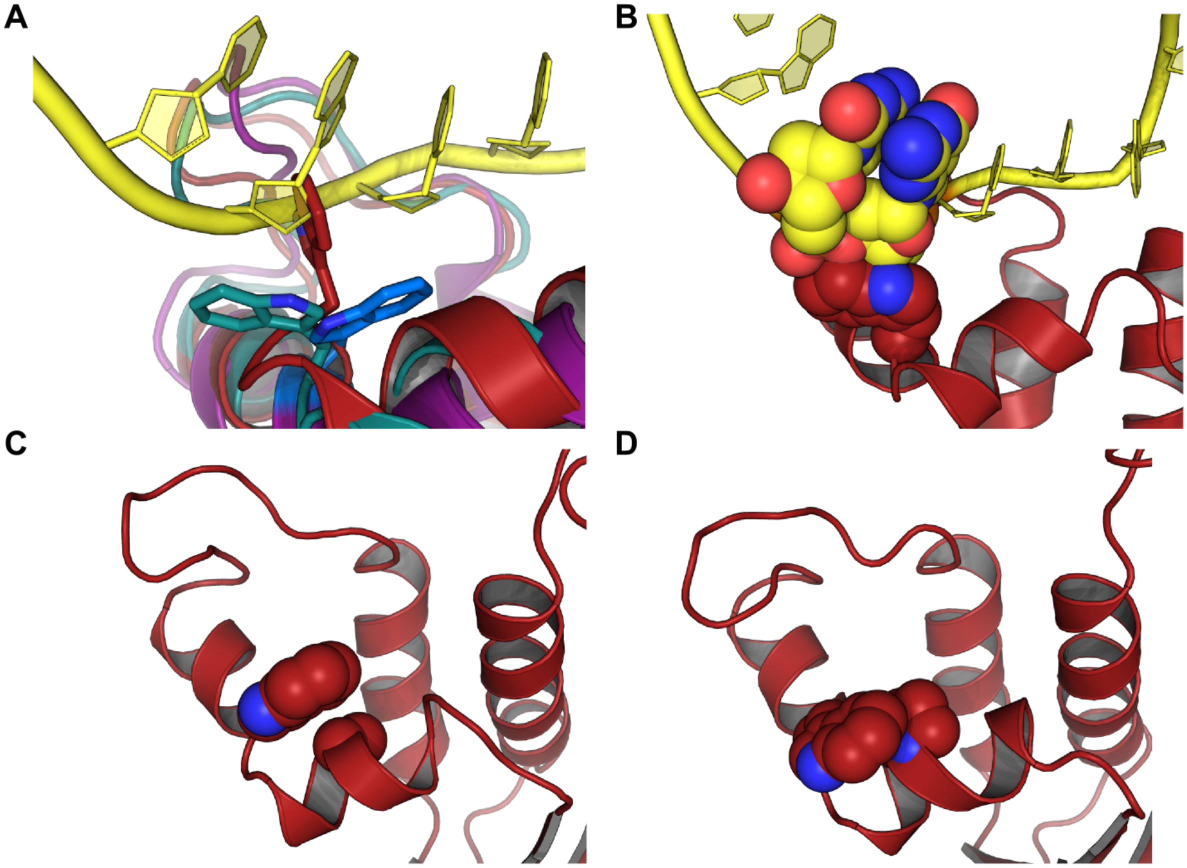

Figure 4.6: **Conformations of the W81 sidechain.** (A) Structural superposition of the ttRNH crystal structure (red), the hsRNH crystal structure (purple), and the alternate W81 conformation sampled in trajectories of wild-type ttRNH (cyan). The χ_2 angles correspond to the *gauche+*, *trans*, and *gauche-* rotamers, respectively. The *gauche+* rotamer clearly creates a steric clash with substrate. (B) The steric environment of the *gauche-* rotamer, which also clashes with the substrate backbone, though less severely than *gauche+*. (C) The packing interface between W81 and its neighbor A77 in wild-type ttRNH. (D) The packing interface between W81 and P77 in the A12S/K75M/A77P trajectories.

The S^2 order parameters of the W81 sidechain N^ϵ group have been measured in ecRNH, ttRNH, and ecRNH iG80b (Table 4.3). The ring is significantly more flexible in ttRNH compared to either of the other two proteins. Remarkably, simulations quantitatively reproduce this pattern [74]. In simulation, ttRNH samples both the *trans* and *gauche-* χ_2 rotamers, while ecRNH trajectories initiated from PDB ID 2RN2 sample only the *trans* conformation and ecRNH iG80b samples only the *gauche-* conformation. Notably, trajectories initiated from the ecRNH structure 1RNH, which occupies the *gauche-* rotamer, predict a low order parameter clearly inconsistent with the experimental data, indicating that this rotamer is not likely to be occupied with significant population in solution. The rotamer distributions from each wild-type simulation are illustrated in Figure 4.7A.

Table 4.3: **S^2 order parameters of W81 ring N^ϵ for various RNase H mutants.**

Protein	Rotamer(s)	S^2 (MD)	S^2 (NMR)
ecRNH WT (2RN2)	*t* (100%)	0.83±0.00[a]	0.83±0.02[b]
ecRNH WT (1RNH)	*g*- (100%)	0.67±0.02	0.83±0.02[b]
ttRNH WT	*t* (93%) /*g*- (7%)	0.66±0.06[a]	0.67±0.02[c]
ttRNH A12S/K75M	*t* (98%) /*g*- (2%)	0.72±0.14	N/A
ttRNH A12S/K75M/A77P 1	*t* (100%)	0.84±0.01	N/A
ttRNH A12S/K75M/A77P 2	*t* (99%) /*g*- (1%)	0.84±0.01	N/A
ttRNH dG80	*t* (100%)	0.80±0.01	N/A
ecRNH iG80b	*g*- (100%)	0.79±0.01[a]	0.82±0.01[c]
ecRNH iG80b Q80L[d]	*t* (97%)	0.67±0.13	N/A
ecRNH iG80b V101R[e]	*t* (48%) /*g*- (52%)	(*g*-) 0.69±0.04 (*t*) 0.85±0.01	N/A
ecRNH iG80b G77A	*t* (66%) /*g*- (34%)	(*g*-) 0.76±0.02 (*t*) 0.82±0.01	N/A

Calculated and experimental NMR S^2 order parameters for the W81 N^ϵ in various RNase H mutants, and the relationship of flexibility to preferred W81 χ_2 rotamer. [a] From N. Trbovic [74] [b] From C.D. Kroenke [239]. [c] From J.A. Butterwick [240]. [d] This simulation was run for 200ns to facilitate sampling of repacking among local hydrophobic groups. [e] This simulation was run for 175ns because a single W81 rotamer transition was observed at 90ns. Note that S^2_{MD} values are calculated in blocks corresponding to the expected global tumbling time of approximately 10ns; therefore rotamer-specific order parameters can be calculated only in cases where the persistence time for each state exceeds 10ns. Populations are not reported for the *gauche+* rotamer, which is quickly abandoned in simulations initiated from the ttRNH crystal structure conformation.

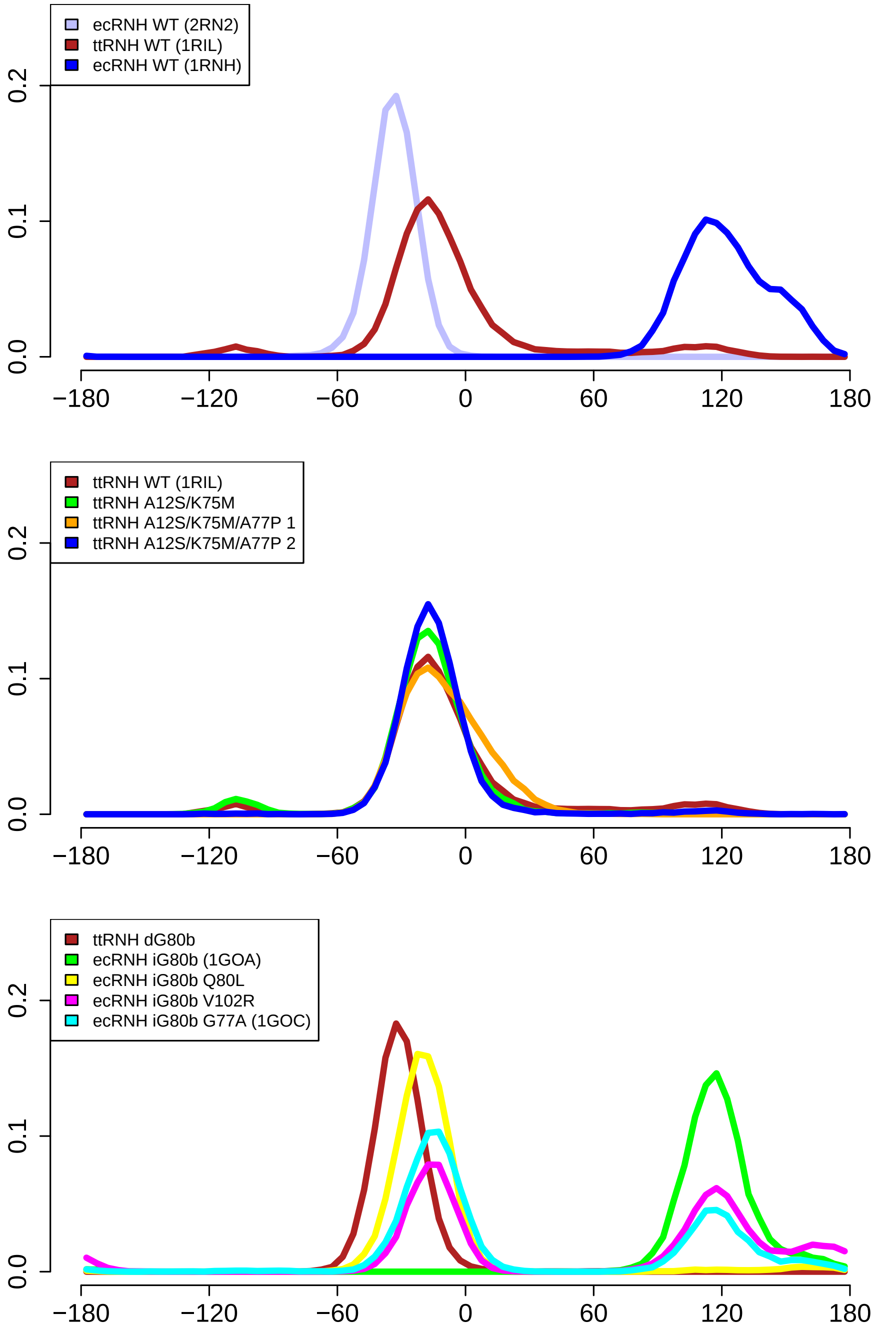

Figure 4.7: χ_2 **distributions of the W81 sidechain.** (Top) The χ_2 distributions for W81 for simulations of the wild-type proteins. (Center) Distributions for activating mutations in ttRNH. (Bottom) Distributions for mutations exploring the coupling between the glycine insertion, position 77, and preferred W81 rotamer populations.

4.4.2 Coupling of W81 to A77

In both A12S/K75M/A77P ttRNH trajectories, the W81 rotamer dynamic is largely abrogated at the 100ns timescale. One trajectory never samples the *gauche-* conformation and the other samples it only once, for less than 2ns. Instead, the addition of the bulky and helix-interrupting proline sidechain at position 77 permits an alteration in local hydrophobic packing in which the indole ring packs against the proline ring, presumably increasing the penalty associated with disrupting this interaction relative to packing against the native alanine residue (Figure 4.6C-D). Reduction of a rotamer population that is sterically incompatible with substrate binding provides a plausible rationalization of the K_m effects of the A77P mutation. The effects of this mutation on k_{cat} are more difficult to interpret, but given that the phosphate backbone of the substrate directly participates in catalysis [241], it is possible that the specific interactions between W81 and substrate contribute to catalytically productive pre-organization of the active site as well as to the initial process of substrate binding.

Interestingly, among bacterial RNase H proteins containing handle loops, those sequences containing a glycine insertion are systematically and dramatically more likely to contain alanine at position 77, while sequences without insertions most commonly contain glycine (Figure 4.9A,C). The simultaneous introduction of these two mutations—iG80b and G77A—into ecRNH has been observed to induce cooperative thermostabilization [136]. The phylogenetic distributions of these two sequence categories suggest that insertions are enriched in Firmicutes (Gram-positive bacteria) while absence of insertion is enriched in Proteobacteria (Gram-negative), although the presence of phylogenetic overlap suggests that the observed sequence pattern is not purely due to common descent (Figure 4.9B,D). (It is worth noting that *E. coli* is a member of the Proteobacteria, and its relatives are likely overrepresented in this dataset.) Only a single sequence in this dataset contains a naturally occurring proline at position 77; this sequence, from *T. equigenitalis* (a microaerophilic proteobacterium), contains N in the insertion position and is 42% identical to ecRNH.

Analysis of more distantly related sequences from all domains of life by the direct-coupling method [221] suggests that position 77 is the second most likely to participate in "evolutionary coupling" interactions, defined as pairs of sequence positions statistically likely to exhibit covariation in large multiple sequence alignments. (The top-ranked position is the i-4 position in helix B, Y73 in ecRNH.) Positions to which residue 77 is coupled are shown in Figure 4.8.

4.4.3 Coupling of W81 to the glycine insertion

Because of the ability of local hydrophobic packing to determine both handle-loop dynamics and W81 sidechain orientation, introduced additional mutations were introduced into ecRNH iG80b intended to "rescue" the protein from its presumably unproductive preferred rotamer by forcing W81 back into the *trans* conformation. The crystal structure of ecRNH iG80b/G77A has been solved and was used to initiate a trajectory for comparison to ecRNH iG80b [136]. Based on the observation that the penalty associated with the *gauche-* conformation can be increased by altering the local hydrophobic packing of the W81 sidechain, the mutation Q80L was introduced to provide a bulkier packing surface. Additionally, based on the observed coupling between V98 and V101 in determining the preferred conformations of the handle loop, the ecRNH iG80b/V101R mutation was tested. All three mutations reduce the population of the presumptively unproductive *gauche-* χ_2 rotamer relative to ecRNH iG80b (Figure 4.7C).

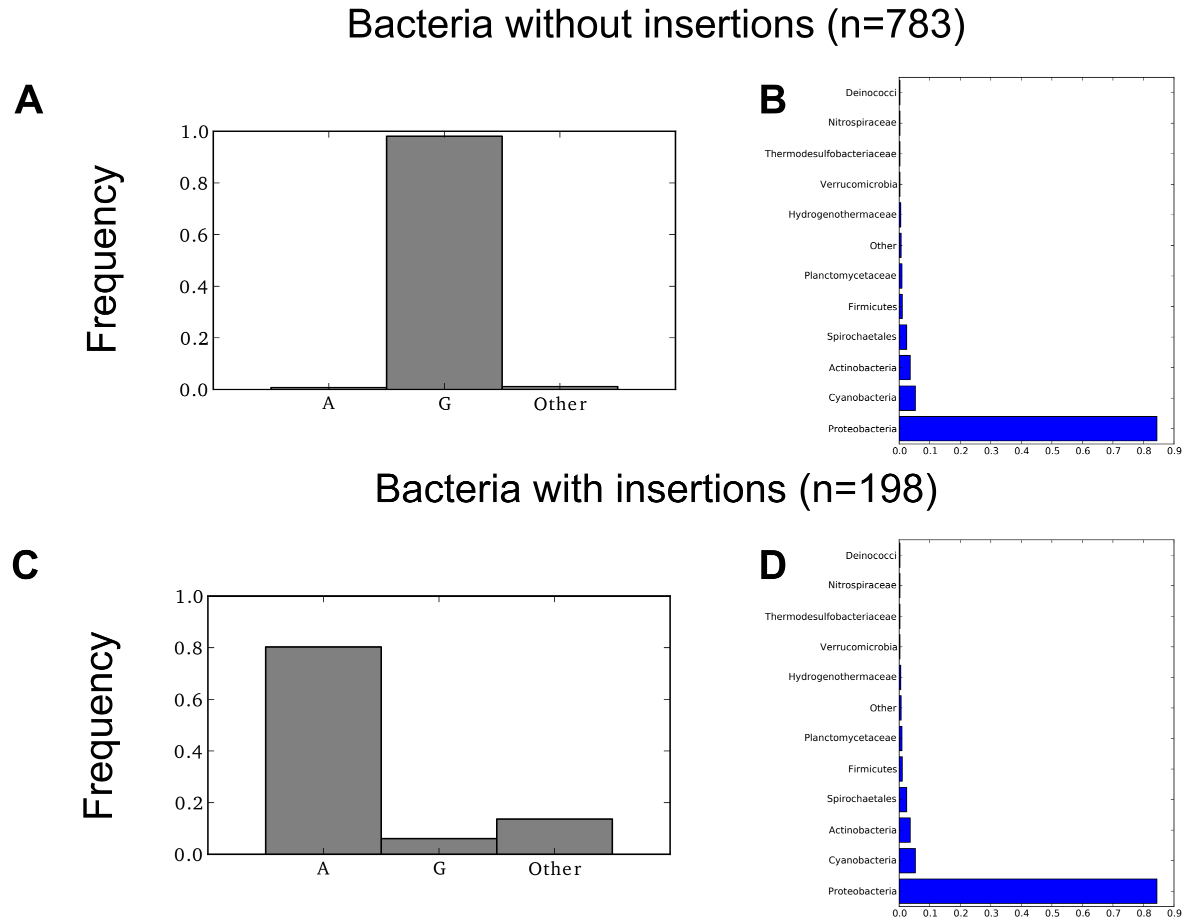

Figure 4.8: **Residue preference and phylogenetic distribution for position 77.** (A) Residues aligned to position 77 (ecRNH numbering) among bacterial sequences that possess handle loops but lack an insertion at position 80b. (B) Phylogenetic distribution of insertion-lacking sequences, dominated by Proteobacteria (Gram-negative). (C) Residues aligned to position 77 among bacterial sequences with handle loops and insertions. (D) Phylogenetic distribution of insertion-lacking sequences, dominated by Firmicutes (Gram-positive). The 'other' residue category contains, in order of frequency in the complete dataset, cysteine (20), serine (12), asparagine (2), proline (1), and threonine (1).

4.5 Backbone dynamics of activating mutants

Interestingly, none of the three trajectories exhibit significantly perturbed average backbone S^2 order parameters relative to wild-type ttRNH, despite changes in the flexibility of particular loops (Table 4.4). Increases in the flexibility of the glycine-rich and active-site

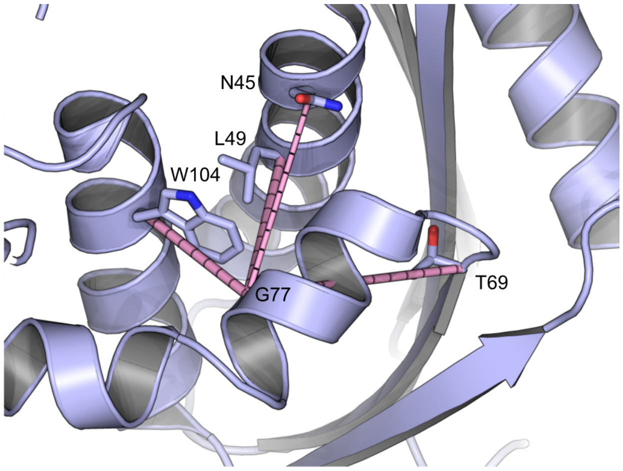

Figure 4.9: **Evolutionary coupling analysis of position 77.** Residues with significant evolutionary coupling to position 77, as determined by the direct-coupling method [221] applied to a multiple sequence alignment covering RNase H homologs in all domains of life assembled by the HHblits tool, mapped onto the structure of ecRNH.

loops, as well as the αA-β4 loop, are offset by increases in rigidity in the handle to produce overall average values that are nearly identical. Although it is not clear how rigidifying the handle (and reducing its population of the open state) is associated with increased catalytic activity, this represents the first observation in this overall simulation dataset of coupling between the behaviors of the active-site loop and the handle.

Table 4.4: **Average S^2 backbone order parameters for ttRNH activating mutants.**

Protein	Mean S^2	Handle S^2	Gly-loop S^2	Active-site loop S^2
ttRNH WT	0.83	0.76	0.86	0.67
A12S/K75M	0.82	0.82	0.84	0.53
A12S/K75M/A77P 1	0.82	0.80	0.84	0.61
A12S/K75M/A77P 2	0.83	0.80	0.84	0.62

Calculated S^2 backbone amide order parameters averaged over the entire protein and selected functionally important loops. The handle is defined as residues 89-100, the gly-loop is defined as residues 12-18 (to include the mutation site), and the active-site loop is defined as residues 121-128.

Chapter 5

Dynamics of the ecRNH active site

5.1 Introduction

Charged and polar residues are overrepresented in the active sites of proteins relative to their overall abundance [242]. In particular, aspartate, asparagine, glutamate, and glutamine contribute functional groups critical for processes such as cofactor binding, acid/base chemistry, and direct participation in enzymatic catalysis. NMR has been extensively used to characterize the dynamics of amino acid sidechains, and has been particularly fruitful when combined with computational approaches—for example, in the study of sidechain motions of arginine [243] and lysine [244], as well as numerous experiments focused on the dynamics of methyl groups [245; 246; 247; 248]. However, surprisingly few experiments have focused on the dynamics of the carboxyl- and carbonyl-containing sidechains [249; 250; 251]

The ribonuclease H active site canonically consists of a highly conserved DEDD motif, represented in the *Escherichia coli* homolog (ecRNH) as D10, E48, D70, and D134 [252] (Figure 5.1A). These residues interact with catalytically required divalent metal ions. Activity has been reported in the presence of Mn^{2+}, as well as inert transition-metal complexes such as cobalt hexaamine, in addition to the physiologically relevant Mg^{2+} [253]. Optimal enzymatic activity occurs at much lower concentration of Mn^{2+} compared to Mg^{2+}, and the presence of very high ion concentration is inhibitory [121]. Measurements of the pKa values of the active-site residues indicate perturbed pKa values for D10 and D70 which normalize upon Mg^{2+} binding, clearly establishing these residue as critical for interaction with ions [205]. The pH optimum for the RNase H reaction *in vitro* is approximately 7.5-8.5 [68], a value at which all active-site residues should be deprotonated [205].

Despite extensive study, the interaction of metal ions with the RNase H active site is poorly understood. Significant differences have been observed between the protein's interactions with Mg^{2+} compared to Mn^{2+}. Co-crystallization studies of ecRNH with high concentrations of Mg^{2+} find a single bound metal ion [120] (Figure 5.1B). By contrast, co-crystallization with Mn^{2+} reveals two bound ions, one in the A site consisting of D10 and D134, and one in the B site consisting of D10, E48, and D70 [123] (Figure 5.1C); the B site is similar but not identical to the previously identified Mg^{2+} site. Single Mn^{2+} sites have been identified in the presence of mutations of E48 and/or D134 [124], both of which are dispensable for Mn^{2+}-dependent activity [254]. Crystallographic studies of related RNases H from the archaeal extremophile *Bacillus halodurans* [118] and the *Homo sapiens* homolog [119] in complex with substrate find two bound ions in the active site (Figure 5.1D).

RNase H domains also occur as a component of the reverse transcriptase protein found in retroviruses and required for viral proliferation [255]. The RNase H domain from HIV has been reported to bind two Mg^{2+} ions even in the absence of substrate [92]; however, recent studies of retroviral RNases H more structurally similar to the known bacterial examples identify a single Mg^{2+} in a position close to that exhibited by the ecRNH structure (Figure 5.1B). Interestingly, there is both structural and dynamic evidence for coupling between the RNase H handle loop and the behavior of the active site. NMR studies of ttRNH in the absence of divalent ions indicate similar dynamic timescales for the handle loop and the active-site loop containing H124 [72], suggesting a coupled motional process. Despite interaction between the positively charged handle loop and the substrate, removing helix C and the handle abolishes Mg^{2+}-dependent but not Mn^{2+}-dependent enzymatic activity in ecRNH [229]. The HIV RNase H subdomain, which lacks helix C and the handle and is inactive, can have its Mn^{2+}-dependent activity restored by insertion of the corresponding sequence from ecRNH [230; 231]. Two structures of the natively handle-containing retroviral RNase H domain from XMRV that both contain a single Mg^{2+} ion differ in its exact location: in the wild-type protein the ion position resembles the Mg^{2+} from ecRNH [212], while in the ΔC mutant the ion resembles the Mn^{2+} B site [210].

The structural diversity of metal-ion locations is summarized in Figure 5.1.

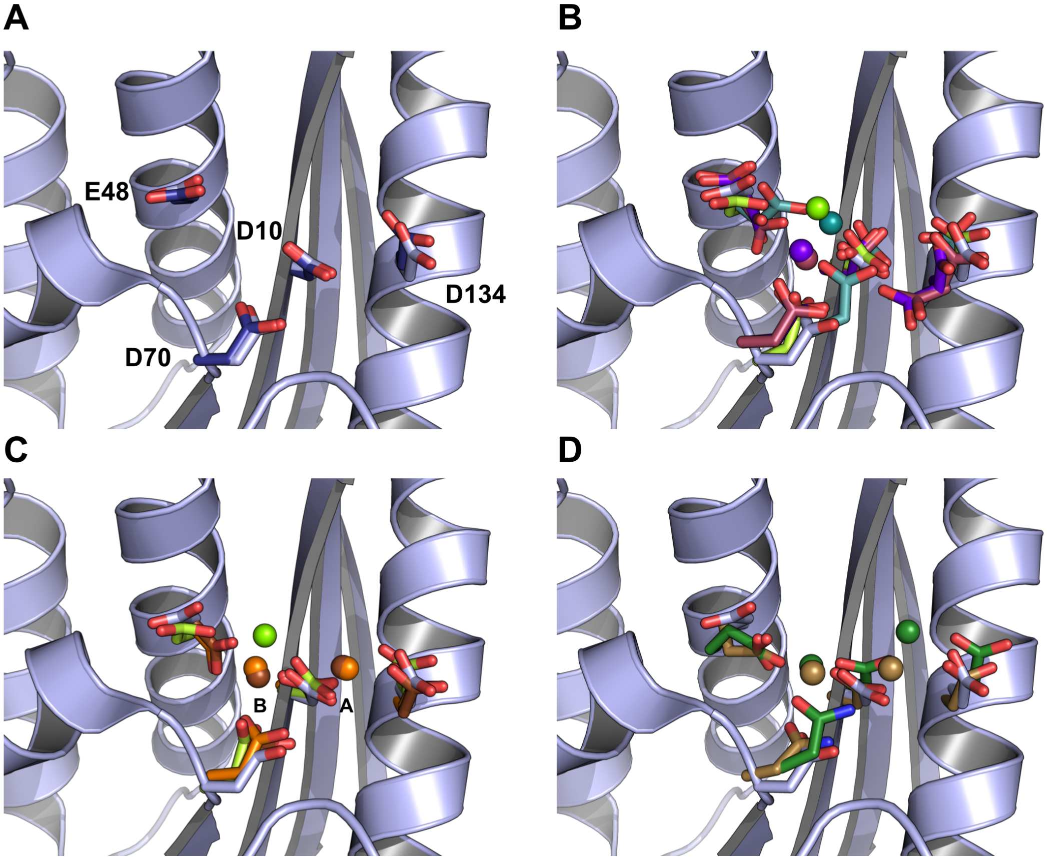

Figure 5.1: **Conformational diversity of metal-ion interactions with ecRNH as determined by crystallography.** In all cases the backbone and active-site sidechains from ecRNH in the absence of ion (PDB ID 2RN2) are shown in light blue for comparison. (A) Structural superposition of the four active-site residues in two structures of ecRNH in the absence of metal ions: 2RN2 (light blue) and 1RNH (dark blue). (B) Structural diversity of RNases H in complex with a single Mg^{2+} ion: ecRNH (1RDD), green; XMRV WT (4E89), dark cyan; XMRV ΔC (3P1G), maroon; MoMLV ΔC (2HB5), purple. The two structures containing helix C and the handle loop (1RDD and 4E89) bind Mg^{2+} in a different position than the two structures of deletion mutants. Additionally, the two deletion mutants both contain two alternate conformations for E48 and D134. (C) Comparison of Mg^{2+} and Mn^{2+} complexes: ecRNH with Mg^{2+} (1RDD), green; ecRNH with Mn^{2+} (1G15), orange; HIV RNase H domain with ecRNH helix C insertion with Mn^{2+} (3HYF), brown. (D) Structural diversity of RNases H in complex with substrate: *Homo sapiens* RNase H with Ca^{2+} ions (2QKK), brown; *Bacillus halodurans* RNase H with Mn^{2+} ions (1ZBI), dark green.

Titration of Mg^{2+} with ecRNH, monitored independently by ^{1}H and $^{25}Mg^{2+}$ NMR, yields a Hill coefficient very close to unity, suggesting that only a single ion binds to the protein in the absence of substrate [253]. The identified binding site has relatively weak affinity; K_d has been reported in the micromolar [253] to low millimolar range [256]. The second site may be occupied only upon binding of substrate [124], possibly due to the presence of high local concentration in the ion cloud of the highly negatively charged nucleic acid molecule. Conformational changes in the active site upon binding the first ion have been suggested as well, with the second site proposed as being responsible for the attenuation of activity at high ion concentrations [121]. Alternatively, a histidine residue (H124 in ecRNH) located in a loop adjacent to the active site has been proposed as responsible for attenuation. Collectively, these results have been used to propose both a one-metal [122; 123; 124] and a two-metal [125; 118; 126] catalytic mechanism. Computational work using the quantum mechanics/molecular mechanics (QM/MM) method applied to the *Bacillus halodurans* [257; 258; 128] and *Homo sapiens* [129] complexes generally supports the two-metal mechanism, with a possible role for a third transiently bound ion [259].

To better understand the dynamics of the RNase H active site upon Mg^{2+} binding, the dynamics of carboxyl- and carbonyl-containing (DENQ) sidechains in molecular dynamics simulations of ecRNH were compared to those inferred from recently developed NMR experiments quantifying motion of these residues at the ps-ns and μs-ms timescales (J.-H. Cho, unpublished data). These results suggest that the active site residues are rigid in the ps-ns timescale while undergoing substantial conformational exchange upon Mg^{2+} binding. This may be interpreted as evidence in favor of electrostatic preorganization for binding the first metal ion, coupled to dynamic reorganization at longer timescales. This work illustrates the utility of combined MD-NMR studies in understanding the dynamic prerequisites for enzymatic catalysis.

5.2 Dynamics of carboxyl- and carbonyl-containing sidechains in the absence of divalent ions

5.2.1 Comparison between MD and NMR data

The experiments conducted monitor the resonances of the carbon atoms directly attached to the carboxyl- and carbonyl-containing functional group, that is $^{13}C^{\gamma}$ for D and N, and $^{13}C^{\delta}$ for E and Q. These studies yielded data of sufficient quality for analysis for 22 out of the 34 DENQ residues in ecRNH. The experimental (S^2_{NMR}) and calculated (S^2_{MD}) order parameters, reflecting motion around the $C^{\beta} - C^{\gamma}$ bonds for D and N and the $C^{\gamma} - C^{\delta}$ bonds for N and Q, are reasonably well correlated (Figure 5.2) with an R^2 value of 0.72, suggesting that the MD simulations can be used to provide atomistic interpretations of the dynamics suggested by the NMR data.

Two out of the four active-site residues—D10 and D134—have higher order parameters in simulation compared to experiment. Simulations were performed for ecRNH protonated to recapitulate preferred protonation states at pH 5.5 (D10 protonated, H124 doubly protonated) and pH8.0 (D10 deprotonated, H124 singly protonated on H^{ϵ}). By contrast, experiments were performed at pD=6.2, which is similar to the experimental pKa of D10 (6.1) [205]. Therefore, the experimental S^2 values for active-site residues should reflect a population average of the protonated and deprotonated states for D10. Inspection of the high-pH simulation reveals that D10 and D134 under these conditions remain rigid due to the formation of a water-mediated hydrogen bond between their sidechains. D134 additionally forms a salt bridge with R138, whose calculated S^2 value is also slightly too high in simulation [243].

Although the experimental data is disproportionately missing solvent-exposed and therefore highly dynamic residues, these residues are well-sampled in MD, which can be used to

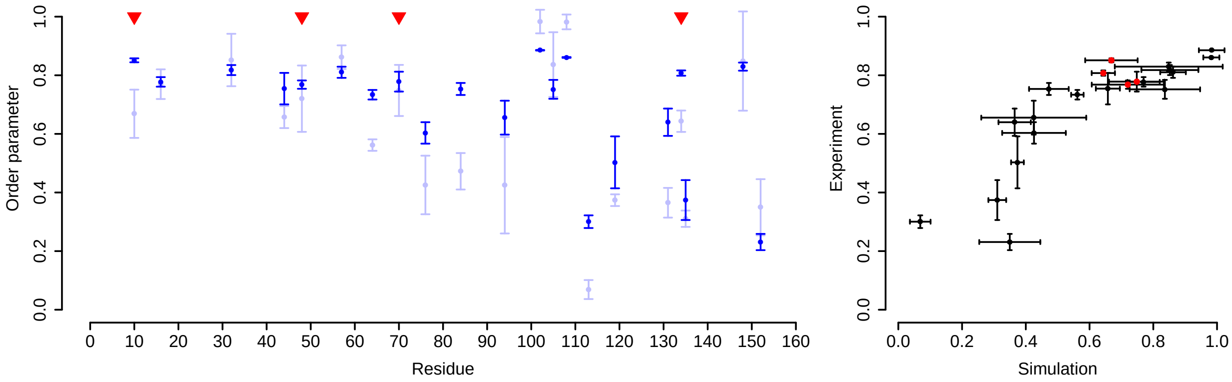

Figure 5.2: **Comparison between experimental and simulated S^2 order parameters for DENQ residue sidechains.** (Left) Experimental (light blue) and simulated (dark blue) S^2 values for the order parameter reflecting motion around the $^{13}C^{\gamma\delta}$ carbon-carbon bond. Experimental values were produced using model-free analysis of the $^{13}C^{\gamma\delta}$ relaxation rates via simultaneous fit of the S^2 and τ_e values; errors of the fits were estimated using Monte Carlo simulations of 300 random data sets (J.-H. Cho, unpublished data). Simulated values were calculated in 10ns blocks to reflect the overall tumbling of the protein; errors reflect the standard error of the mean. Simulated values were scaled by 0.93 to account for differences in bond-length assumptions. Active-site residues are indicated as red triangles. (Right) Correlation between experimental and simulated values. Active-site residues are indicated in red. The Pearson's R^2 value for the 21 residues analyzable by the model-free formalism is 0.72.

obtain unbiased statistics regarding the flexibility of each residue type. Overall, both the MD and NMR data imply that the shorter-sidechain D and N residues are collectively more rigid than are E and Q (Table 5.1). This is consistent with previous observations on calbindin D_{9k}, the only other protein for which such data has been measured [250]. However, MD may also systematically overestimate the rigidity of the shorter sidechains, an observation that is only moderately improved between the 99SB and 99SB-ILDN force fields, which have average S^2 values for N residues of 0.79 and 0.73 respectively. (Interestingly, this is the opposite of the pattern that has been observed for backbone amide S^2 values, which tend to be systematically underestimated due to population of spurious conformations not likely to be represented in solution [236].) Nevertheless, the relative rigidity of all

Table 5.1: **Average S^2 values per sidechain type for DENQ residues.**

Residue type	MD	NMR
D	0.81	0.76
E	0.60	0.58
N	0.79	0.63
Q	0.43	0.42

Comparison of average S^2 values per sidechain type as determined by MD and NMR.

four sidechain types measured by NMR is reproduced in simulation. These data support the hypothesis based on structural bioinformatics that D is preferred in enzyme active sites due to its increased rigidity [242].

5.2.2 Effects of variations in simulation conditions

Variations in simulation conditions produce S^2_{MD} values for active-site residues that are more similar to each other than to S^2_{NMR} values, indicating that the details of the simulation setup do not determine the quality of experimental agreement (Figure 5.3). The only changes in simulation conditions that produce substantial differences are the use of the AMBER99SB-ILDN force field, which contains modifications to the sidechain parameters for D and N residues, and the use of the TIP4P water model instead of the more computationally efficient TIP3P. Two solvent-exposed residues—N16 and N84—exhibit differences between 99SB-ILDN and 99SB, and N84 and N100 differ between TIP3P and TIP4P; however, the dynamics of the active-site residues are not affected. Importantly, additional sampling time did not substantially change S^2, as demonstrated by comparing a 100ns ecRNH trajectory to a 1μs trajectory; the ps-ns timescale dynamics of the DENQ sidechains are therefore well sampled at 100ns of simulation time. It is likely that discrepancies in S^2 values for the active-site residues arise not from incomplete conformational averaging alone, but instead from the inability of standard molecular mechanics force fields to model changes in protonation state.

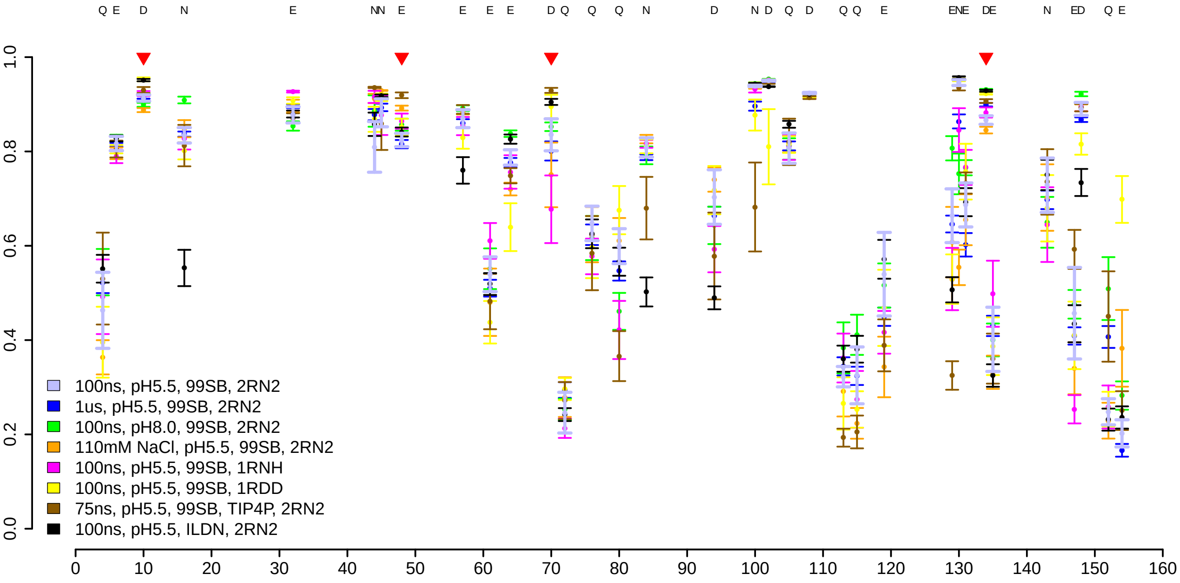

Figure 5.3: **Variation in calculated DENQ sidechain order parameters with simulation conditions.** Calculated S^2 values are shown for various simulation conditions with standard errors of the mean. The four active-site residues are indicated with red triangles. In no case does the difference between any two simulations reach statistical significance.

5.3 Structural determinants of ps-ns timescale sidechain dynamics

Sidechain S^2 values are at best weakly correlated with backbone amide S^2 (R^2=0.32), relative surface area in the crystal structure (R^2=0.30), MD-averaged solvent-exposed surface area (R^2=0.38), or (for D and E residues) with the size of the measured pKa perturbation (R^2=0.03) (Figure 5.4. There is a correlation with the crystallographic B-factors from the ecRNH structure (PDB ID 2RN2), despite the differences in origins of the motions measured by each parameter (R^2=0.57). The weak correlation with solvent exposure is in contrast to observations of the analogous S^2 value of the amino group of lysine sidechains in ubiquitin [244] and N^ϵ of the guanidinium group of arginine sidechains in ecRNH [243], which have been observed to be correlated to solvent accessibility.

Both sidechain flexibility and ΔpKa show a similar pattern in their relationships to RSA (Figure 5.5). Buried residues tend to be rigid, but solvent-exposed residues exhibit highly heterogeneous dynamic behavior. Similarly, exposed residues tend to have unperturbed pKa values, but partially buried residues vary significantly. Sidechains with similar dynamic behavior tend to colocalize within the structure; highly rigid sidechains cluster within the active site and in the interfaces among helices A, C, and D (Figure 5.6). Other than D10 and D70, carboxylate-containing residues with perturbed pKa values tend to be those that participate in highly persistent salt bridges.

5.4 Active-site dynamics in other RNase H homologs

Given that all known RNase H homologs have extremely similar active-site structures, it is likely that measurements made on the ecRNH protein can be generalized to other RNase H homologs, at least those that contain the helix C/handle loop structure known to be

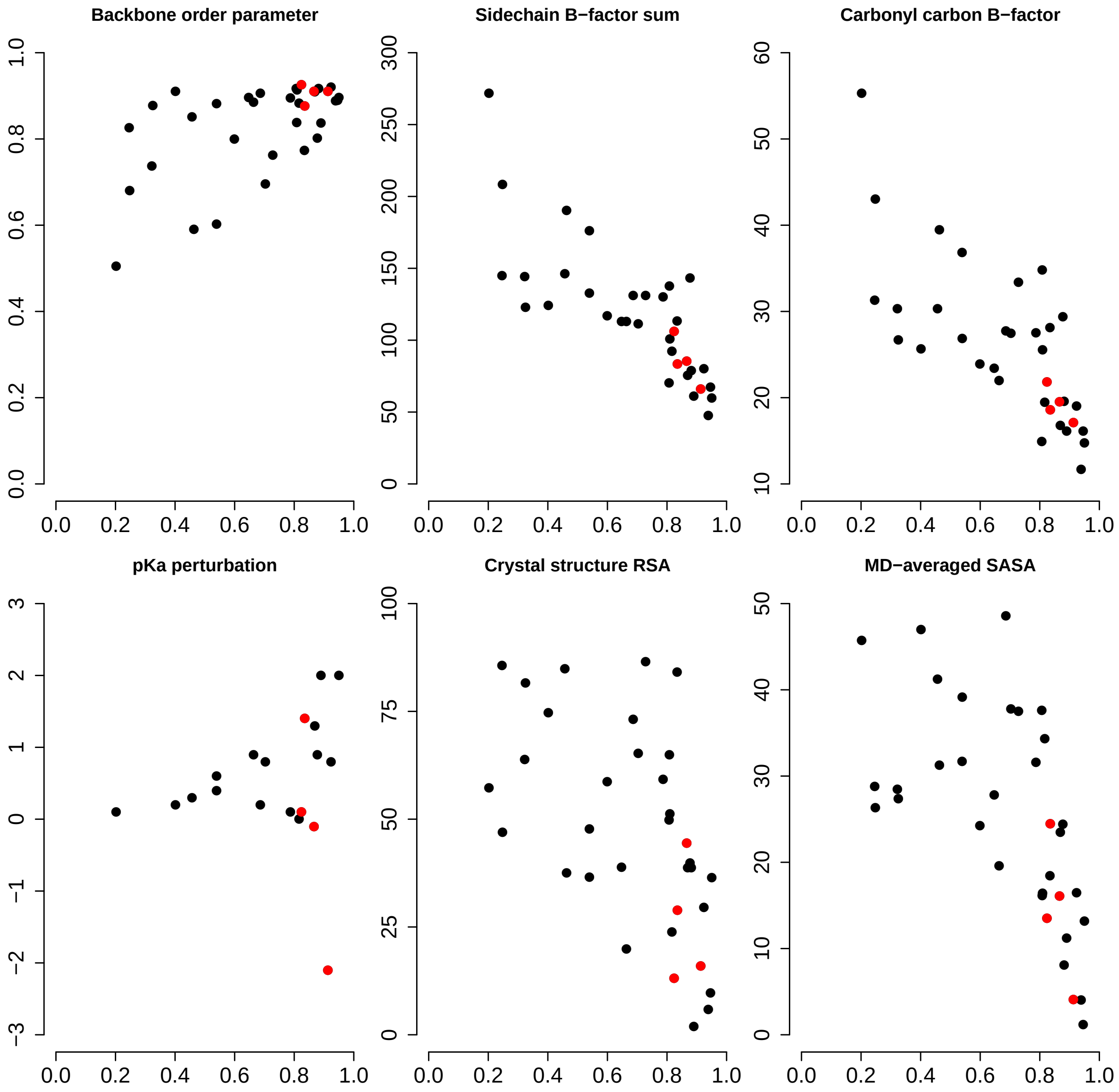

Figure 5.4: **Correlation of DENQ sidechain order parameters with various structural parameters.** Calculated S^2 values for DENQ sidechains are correlated to various structural parameters expected to influence or be influenced by local sidechain dynamics. The four active-site residues are shown in red. Crystal structure RSA was calculated using GETAREA [260]. MD-averaged SASA was calculated using VMD [261]. To avoid biasing statistics due to omission of highly flexible residues, all correlations use the calculated rather than experimental sidechain S^2 values.

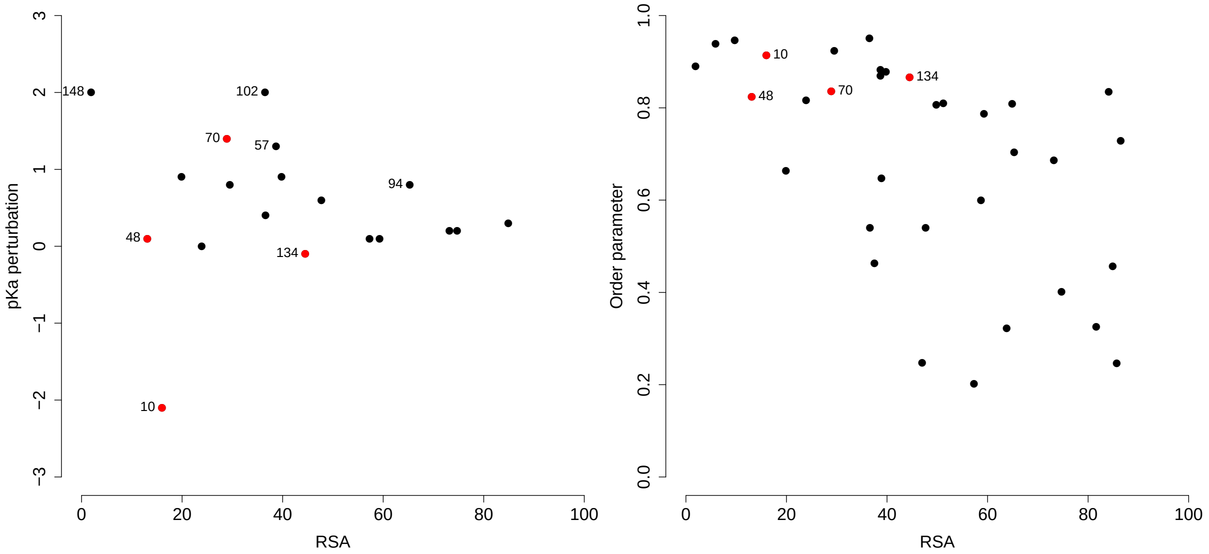

Figure 5.5: **Effects of buried surface area on dynamic behavior for carboxylate-containing sidechains.** (Left) Perturbation in pKa of each carboxylate sidechain as a function of the relative surface area (RSA) calculated from the crystal structure 2RN2. (Right) Calculated S^2 order parameter as a function of RSA. Active-site residues are shown in red. The residue numbers for significant outliers from the pKa trend are shown.

coupled to metal-binding behavior. S^2_{MD} values were therefore calculated for the remaining bacterial RNase H homologs of known structure, as well as for hsRNH in the absence of substrate.

As might be expected from the high level of structural conservation in the active-site region, the five handle-region-containing RNase H homologs compared differ very little in the dynamics of their active site residues (Figure 5.7). Notably, the trajectory initiated from the hsRNH structure, which was solved in the presence of substrate and which contained a Na^+ ion in a position similar to the B site in ecRNH, differs very little from trajectories initiated from any other RNase H structure lacking these additional components. This observation provides strong support for the interpretation that the rigid active-site residues are electrostatically preorganized for metal-ion interactions even in the unbound state.

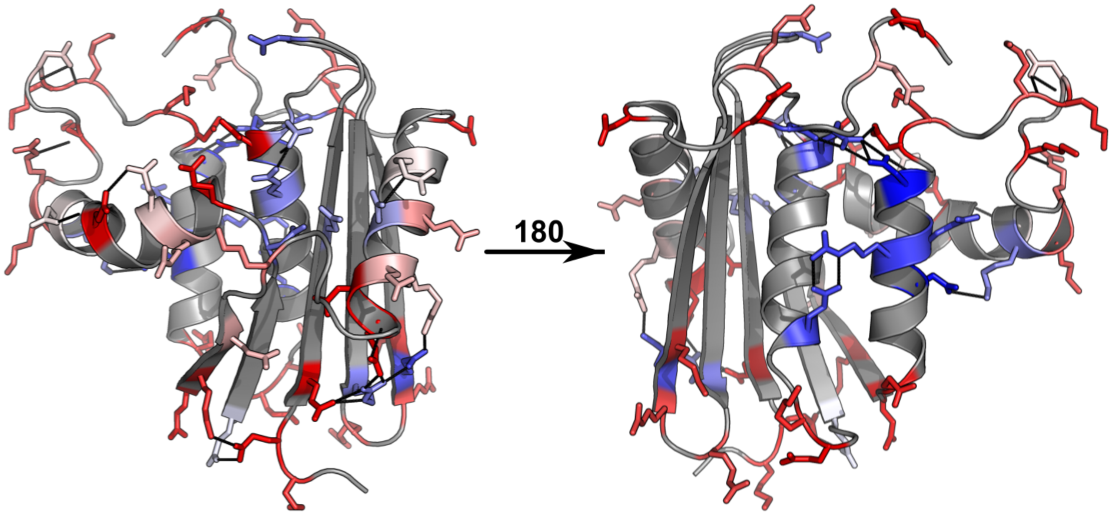

Figure 5.6: **Colocalization of sidechain dynamic behavior in ecRNH.** Residues are colored by S^2 values of their sidechains, with red indicating flexibility and blue indicating rigidity. Experimental values for arginine residues were reproduced from [243]. Because measurements of the lysine residues have not been made, and require much lower temperatures than those at which other residues have been studied [244], simulated S^2 values were used instead.

In order to better understand the dynamic processes that might lead to the adoption of the "Mg^{2+}-like" binding conformation assumed by ecRNH, rather than binding in the Mn^{2+} B site, additional simulations in the absence of divalent ions were performed on a set of retroviral RNase H homologs chosen to directly compare the behavior of proteins that do and do not contain helix C and the handle loop. As detailed in Figure 5.1, differences are expected in the ion-binding behavior of homologs with and without this sequence. These results are summarized in Figure 5.7B. In brief, no significant differences are observed between simulations initiated from the XMRV full-length structure compared to the ΔC mutant, between the XMRV ΔC mutant compared to the HIV homolog (which naturally lacks this sequence), or between any of the retroviral domains compared to ecRNH. This result suggests that the preorganization of the active site on the ps-ns timescale is not

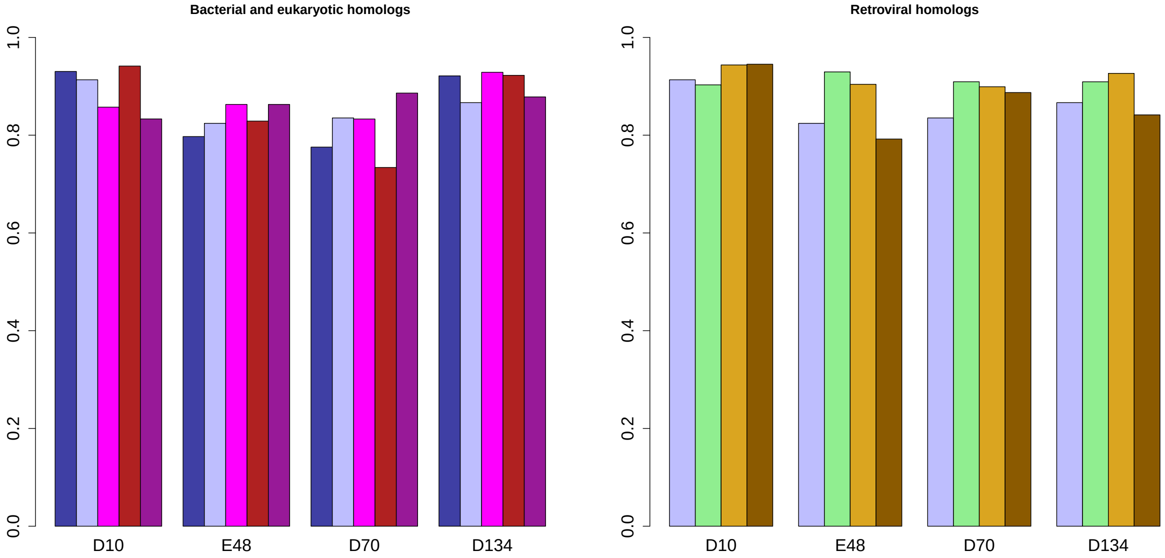

Figure 5.7: **Active-site dynamics in RNase H homologs.** Calculated S^2 values are shown for the four active-site residues in RNase H homologs. Left: soRNH (dark blue), ecRNH (light blue), ctRNH (magenta), ttRNH (red), hsRNH (purple). Right: ecRNH (light blue), XMRV WT (green), XMRV ΔC (yellow), HIV (brown). All simulations were carried out at 300K in the AMBER99SB force field with TIP3P water with structures protonated to reflect a pH of 5.5.

modulated by differences in amino acid sequence, but rather is inherently imposed by the overall protein fold. Although crystallographic evidence suggests there is a difference in ion binding between proteins that contain a handle and those that do not, this difference does not appear to be reflected in the apo-state dynamics of the proteins at the ps-ns timescale. Simulations of the *Homo sapiens* RNase H homolog and the *Thermus thermophilus* argonaute protein (a distant RNase H homolog with similar active-site architecture) in complex with two Mg^{2+} ions and phosphonate-based substrate analogs find the active site to be rigid under these conditions as well; in conjunction with the present data, these results imply that active-site preorganization is a general property of this larger family of nucleases [262]. Although selective inhibitors of the HIV RNase H domain have been developed based on the hypothesis that the metal ion dependence of the HIV domain's catalytic mechanism differs from that of the human homolog [125], it is likely that this selectivity derives from differences in the relative affinity of the two metal ion binding sites—perhaps due to differences in reorganization after binding of the first ion—rather than differences in mechanism *per se*.

5.5 Conformational consequences of Mg^{2+} binding

5.5.1 Experimental observations of R_{ex} accompanying Mg^{2+} binding

Experimental investigations additionally characterized the effect of Mg^{2+} on the dynamics of the DENQ sidechains. These results are summarized in Figure 5.8. Chemical exchange line broadening that can be attributed to the presence of Mg^{2+} (R_{ex}) is primarily, though not exclusively, observed in residues in or near the active site. Assuming a motional process that is fast on the NMR timescale, all residues involved in the same process should be collinear on a plot of the square of the observed chemical shift difference $\Delta\omega^2$ between the apo and Mg^{2+}-bound forms, versus the measured R_{ex} value. Figure 5.8B shows that two distinct processes are observed; residues can be assigned to two groups based on their observed perturbations. Group 1 consists of N44, E57, and D102. Group 2 consists of E48, D10, D70, and D134 (although D10 is too broadened in the Mg^{2+}-bound state to observe, it has been included in group 2 on the basis of its extremely large chemical shift difference). Group 2 is thus coextensive with the catalytic carboxylate residues. The positions of these residues are indicated in Figrue 5.9.

As a result, these data do not distinguish between the multiple candidate binding modes for the interaction of Mg^{2+} ions with the active-site residues. However, the group 1 residues suggest that perturbations due to metal binding are experienced by residues in the helix A-helix D interface as well as those in the active site. Both E57 and D102 are involved in salt bridge interactions with arginine residues that form part of the dimeric coiled-coil-like interface between these two helices. These arginine residues have previously been identified as having highly rigid guanidinium groups on the ps-ns timescale in both simulation and experiment [243]. Here their carboxylate salt-bridge partners are identified as similarly

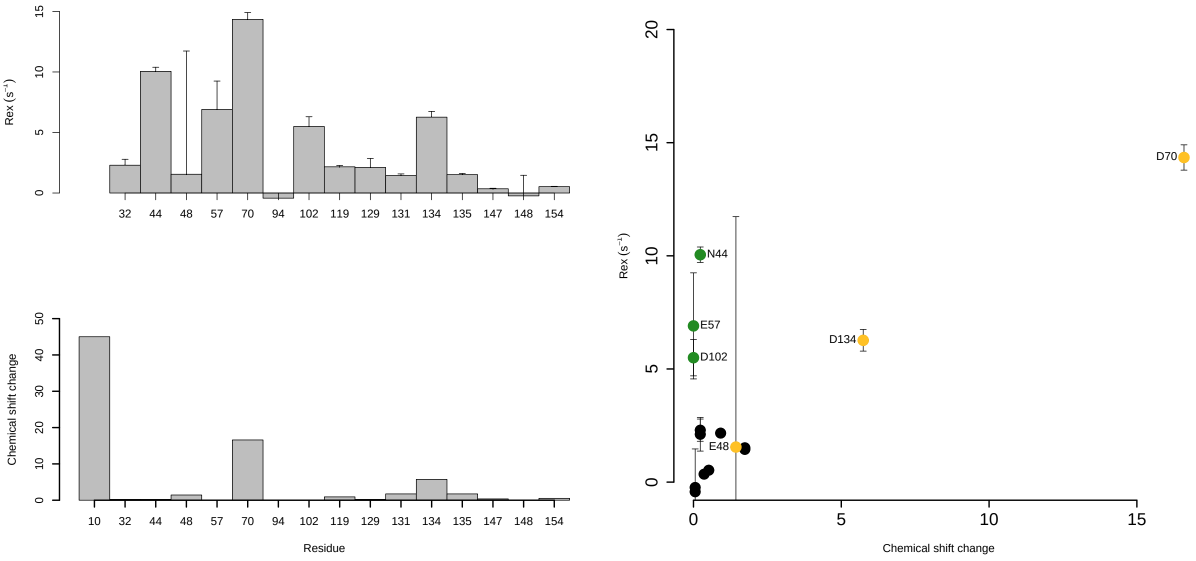

Figure 5.8: **Experimental measurements of** Mg^{2+} **binding to the active site of ecRNH.** (A) Contribution of Mg^{2+} to chemical exchange line broadening (R_{ex}). D10 was too broadened to observe. (B) Squared change in chemical shift ($\Delta\omega^2$) of each $C^{\delta\gamma}$ in the presence of Mg^{2+}. (C) R_{ex} as a function of $\Delta\omega^2$, illustrating the two groups into which the DENQ sidechains can be clustered. Group 1 (green) shows little $\Delta\omega^2$, while group 2 (yellow) shows large $\Delta\omega^2$ and encompasses the active-site residues. (J.-H. Cho, unpublished data.)

rigid on a short timescale, but sensitive to perturbations originating from the active site. Previous work studying perturbations to the backbone upon ion binding did not identify any changes in this region [256], suggesting that if dynamics are transduced from the active site to the helical interface, it occurs through the behavior of the sidechains. Alternatively, the observation of disruption of salt bridges upon increase in ionic strength of the buffer may simply be due to competition between Mg^{2+} and arginine for interaction with the charged carboxylates.

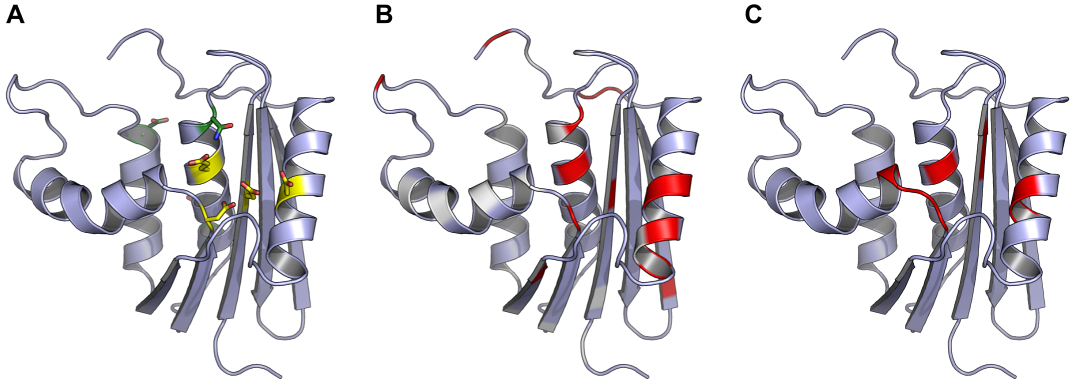

Figure 5.9: **Experimental measurements of the effects of** Mg^{2+} **binding mapped onto the structure of ecRNH.** (A) Group 1 (green) and group 2 (yellow) residues as defined by measurements of R_{ex} and $\Delta\omega^2$ (Figure 5.8). (B) $\Delta\omega$ values for sidechain $C^{\gamma\delta}$ reflecting perturbation due to Mg^{2+}. White corresponds to no chemical shift change, red corresponds to a large change, and non-DENQ residues are shown in light blue. (C) Residues previously shown to experience backbone ^{15}N or ^{1}H chemical shift changes upon binding Mg^{2+} [263].

5.5.2 Simulations of ecRNH in complex with Mg^{2+}

In order to more fully understand the dynamics of ecRNH in the Mg^{2+}-bound state, simulations were conducted of ecRNH in the presence of a single Mg^{2+} ion initiated from the crystal structure of ecRNH in the Mg^{2+}-bound state (PDB ID 1RDD) [120]. The Aqvist parameter set was used for the Mg^{2+} ion [216]. However, the position of the ion identified in this structure is not stable in simulation and exits the binding site immediately upon initiation of the trajectory. The ion transiently interacts with the protein at a variety of sites on the protein surface over the course of the 89ns trajectory but never returns to its original position in the active site (Figure 5.10A).

Historically, simulation of the behavior of multivalent ions using standard molecular mechanics force fields has been a long-standing challenge. Because of the quadrupolar effects of high local charge density, multivalent ions are not well described by the assumption that interatomic interactions can be decomposed into a series of pairwise sums. It is therefore possible that the instability of this position in simulation is an artifact of force field errors. However, given that ions in this position are not observed in the substrate-bound structures of RNase H homologs (Figure 5.1D), and that the B-factor of the Mg^{2+} ion in the 1RDD structure is much higher than those of the surrounding residues (Figure 5.10B), it is also possible that this position does not reflect the most stable conformation of the protein-ion complex in solution.

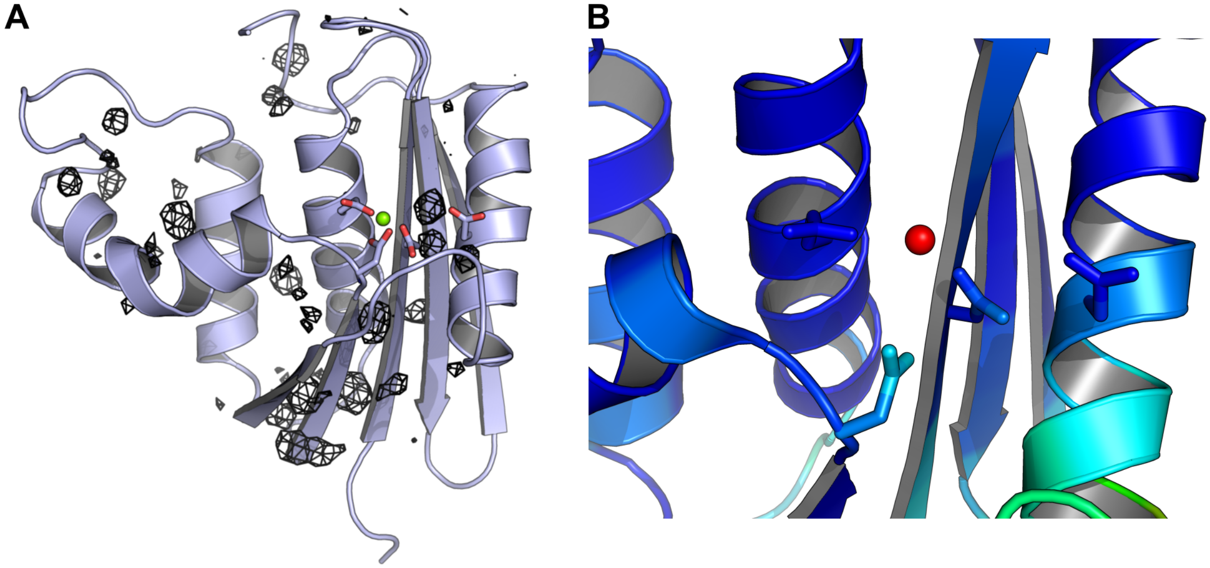

Figure 5.10: Mg^{2+} **ion dynamics in a simulation initiated from the ecRNH** Mg^{2+}**-bound crystal structure.** (A) Occupancy map from an 89ns simulation initiated from the ecRNH structure solved in the presence of Mg^{2+} (PDB ID 1RDD), contoured to 0.05% occupancy (corresponding to at least 45ps total residence time). The ion exits the active site and interacts with a variety of regions on the protein surface. (B) The active-site region of the 1RDD structure, colored by atomic B-factor. The B-factor of the ion is substantially larger than the surrounding residues, and is in fact larger than the B-factor of any other atom in the structure save crystallographic waters.

Additional simulations were carried out under the same conditions for single Mg^{2+} ions in each of the two Mn^{2+} binding sites identified for ecRNH. Because the crystal structure of ecRNH solved in the presence of Mn^{2+} (PDB ID 1G15) exhibits disorder in both the active-site and handle loops [123], the ion positions were instead modeled into the apo ecRNH structure (PDB ID 2RN2) by superposition. For the model of the B-site Mg^{2+} ion, the rotamer of E48 was also corrected to match that observed in the 1G15 structure. For comparison to an alternative homolog, the Mg^{2+} ion in the B site was also modeled into ttRNH (PDB ID 1RIL), whose structure was also solved in the absence of divalent ions.

Mg^{2+} ions were found to be stably associated with the ecRNH active site in both simulations, despite the fact that the ions were modeled into a structure that did not originally contain them (Figure 5.11). This observation clearly supports the hypothesis that electrostatic preorganization in the active site promotes ion binding. It is possible that the effectiveness of this modeling procedure was facilitated by a well-documented feature of crystal packing in ecRNH, in which the amino group of a lysine sidechain in a neighboring molecule inserts into the negatively charged active site in a position approximating the B site. However, a short simulation of ttRNH, whose structure does not contain this contact, with Mg^{2+} modeled into the B site was also stable, suggesting that crystal contacts in ecRNH are not responsible for the observation of preorganization in its active site.

The presence of Mg^{2+} located in either the A or the B site did not substantially affect the dynamics of the active-site residues. All four residues remain highly rigid in the presence of an Mg^{2+} in either position (Figure 5.12). The major difference between the unbound, A site, and B site trajectories' carbonyl-sidechain dynamics was observed in a short loop between helix D and β-sheet 5. This region is significantly heterogeneous in other variations of simulation conditions (Figure 5.3), suggesting that it is simply incompletely sampled at the 100ns timescale rather than significantly perturbed by ion binding. No significant differences in the behavior of these residues are observed experimentally at the μs-ms timescale.

Of the four conserved catalytic residues, D134 is known to be somewhat dispensable; catalytic activity is retained, though reduced, by substitutions with N or H, which also

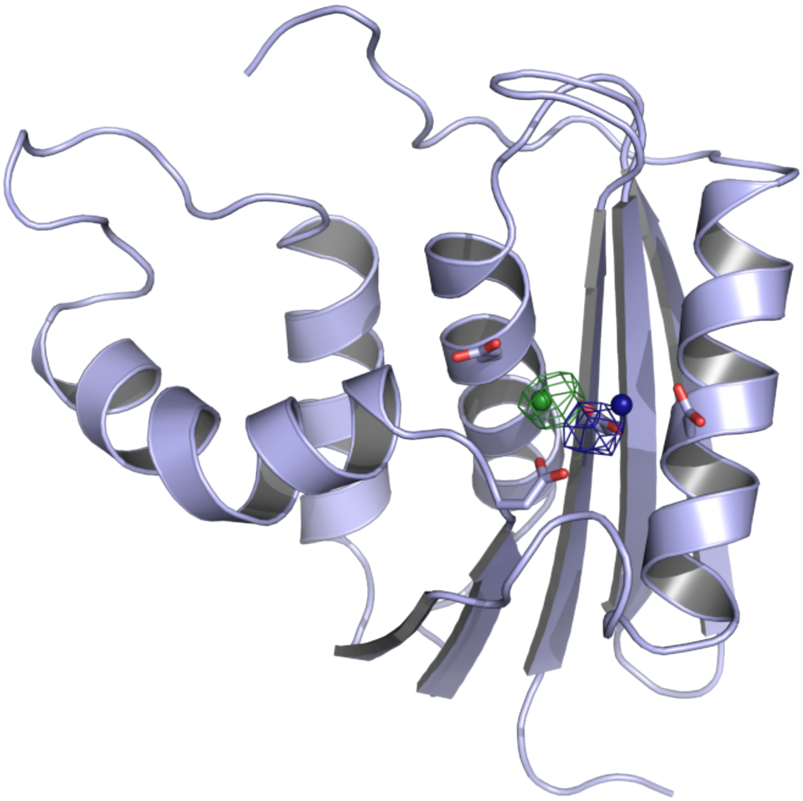

Figure 5.11: **Occupancy maps for** Mg^{2+} **positions for ecRNH trajectories with single ions in the A and B sites.** Maps contoured at 10% occupancy for the A site (blue) and B site (green). The conformation of the four active-site residues in apo ecRNH and the positions of the two Mn^{2+} from which the trajectories were initiated are shown for comparison. Drift from the original A site position is clearly shown.

increase thermostability. In conjunction with crystallographic evidence, this suggests that the B site is the one occupied in the absence of substrate. Because measurements of the sidechain $^{13}C^{\gamma\delta}$ resonances by NMR could not clearly distinguish the behavior of D134 (the unique participant in the A site) from E48 (the unique participant in the B site), comparisons of the two trajectories provide an additional opportunity to distinguish between these two sites.

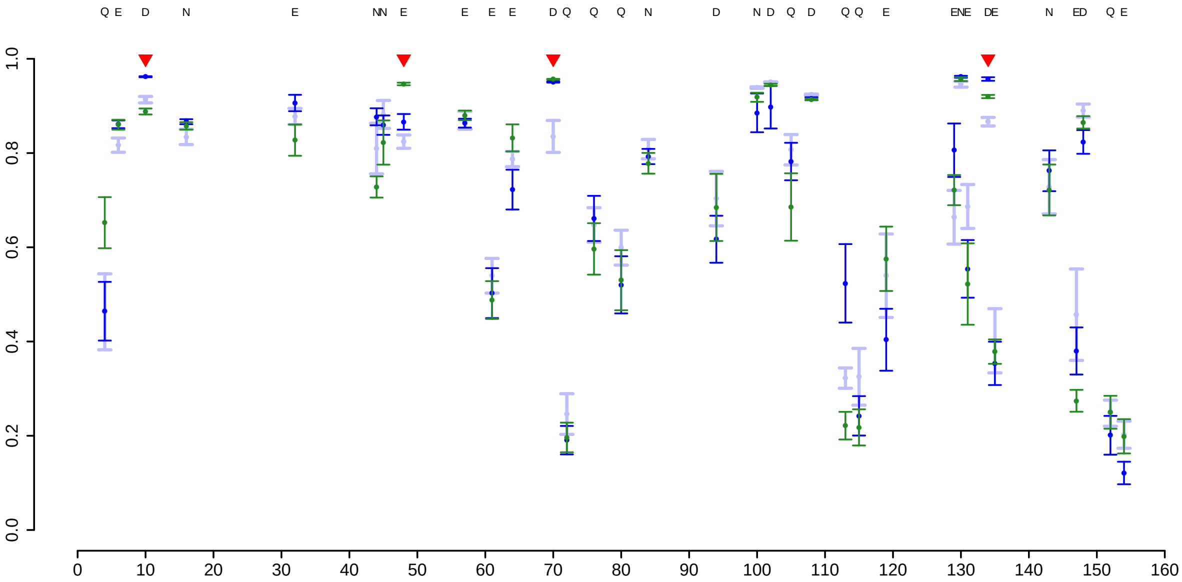

Figure 5.12: **Differences in DENQ sidechain S^2 values in the presence of** Mg^{2+} **ions in ecRNH.** Calculated S^2 values are shown for various simulation conditions with standard errors of the mean: ecRNH apo (light blue), A site (dark blue), B site (green). The four active-site residues are indicated with red triangles. In no case does the difference between any two simulations reach statistical significance.

Although single metal ions in both sites were stably bound to the protein, the RMSD over the course of each 100ns trajectory was larger for the ion in the A site (1.2Å) compared to the B site (0.6Å), which in turn is similar to the RMSD of a 30ns control simulation of ttRNH with an ion modeled into the B site (0.6Å). Additionally, a small amount of motion in the direction of the B site was observed for this ion; the initial and final positions differ by 1.7Å (Figure 5.11). (By comparison, the A and B sites are about 4Å apart.)

Distinct conformations were also observed for several neighboring residues, reflecting reorganization of local hydrogen bonding networks to accommodate ion binding in each of the two sites. N45 does not differ significantly in sidechain rigidity between the two trajectories, but it does differ in conformation: in the A site trajectory, it is oriented away from the substrate-binding site and participates in a network of interactions that also includes the conserved site T43, while in the B site N45 is oriented into solvent and occupies the rotamer found in the hsRNH-substrate complex.

The hydrogen-bonding network surrounding D134 unsurprisingly differs considerably between the A and B site trajectories. Occupancy of inter-sidechain hydrogen bonds in this region is summarized in Table 5.2. H124, which interacts with substrate in the hsRNH complex and is known to be associated with product release, forms hydrogen bonds with D134 in in the B site trajectory, partially displacing one of the hydrogen bonds formed between D134 and R138 in the apo trajectory. By contrast, H124 interacts primarily with E131 in the A site trajectory, while D134 coordinates Mg^{2+} in a monodentate manner, partially displacing the R138-D134 interaction. This conformation too is at odds with experimental evidence, since E131 experiences minimal chemical shift perturbation upon Mg^{2+} binding.

The salt bridge between D134 and R138 has potential functional consequences. R138 is conserved in hsRNH, but adopts a distinct conformation in which it is oriented toward the backbone of the RNA strand of the substrate and forms a hydrogen bond to the phosphate group of the nucleotide adjacent to the scissile phosphate. The maintenance of a strong hydrogen-bonding interaction in the apo state thus minimizes the entropic cost of binding substrate (Figure 5.13).

These results collectively add to prior experimental evidence that the B site is the primary site for metal ion binding in the absence of substrate. Furthermore, the presence of a metal ion in the B site may induce reorganization of the surrounding sidechains into conformations conducive to subsequent substrate binding.

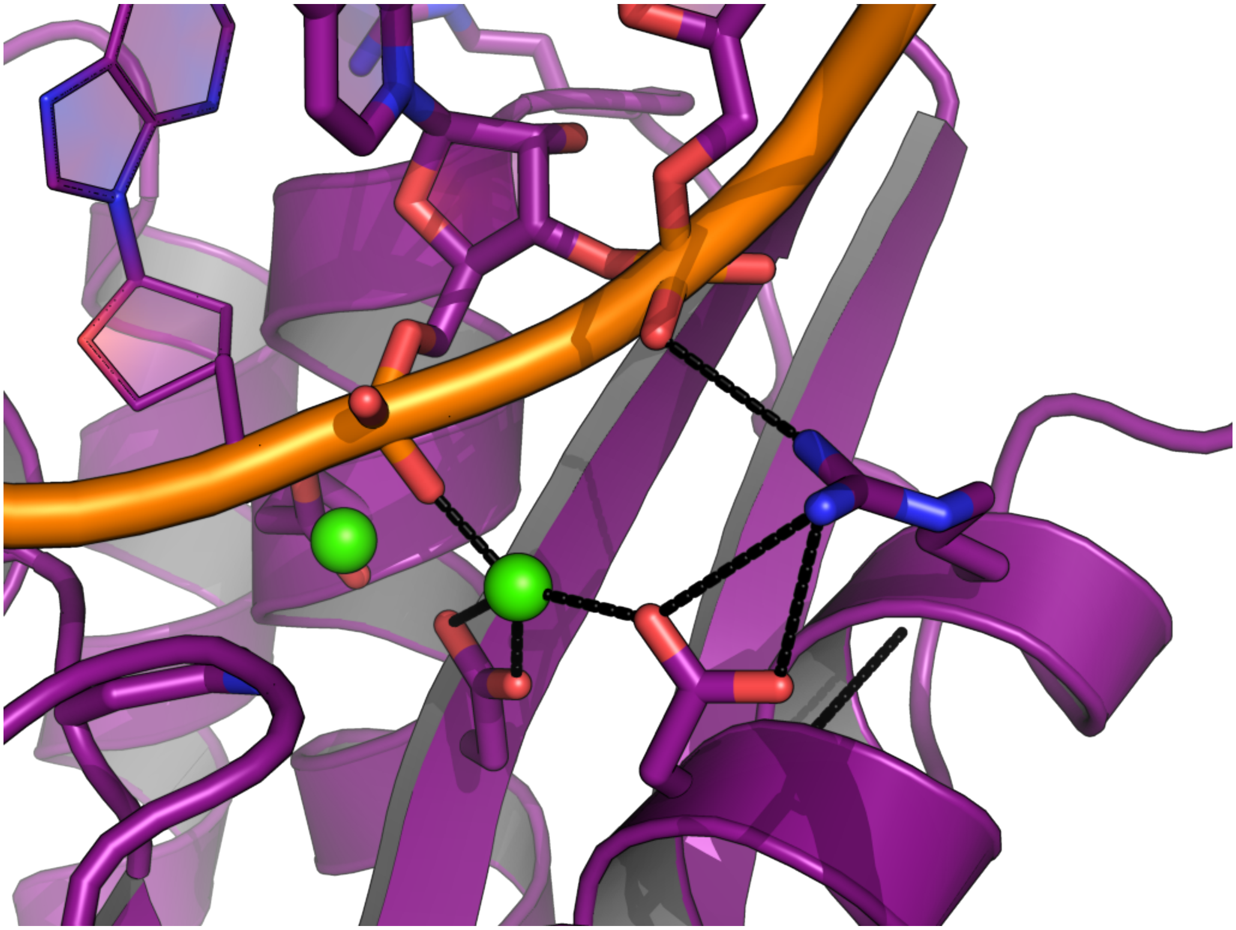

Figure 5.13: **Hydrogen-bonding environment of the A site of hsRNH in complex with substrate.** Interactions between D134, R138, metal ion A, and the phosphate backbone of the RNA strand are shown. The preorganization of the D134-R138 salt bridge in the apo state of ecRNH likely minimizes the entropic cost of forming this interaction upon substrate binding.

Table 5.2: **Hydrogen bond occupancy in the network surrounding D134 in ecRNH.**

H-bond	Apo	A Site	B Site
H124-E131	2.2%	47.0%	16.7%
H124-D134	0.8%	1.5%	23.8%
H124-E135	0%	0%	0.3%
R138-E131	0%	11.3%	0%
R138-D134	74.1%	48.4%	65.5%
R138-E135	43.4%	29.8%	50.1%

Comparison of hydrogen-bonding environments in the network surrounding the active-site residue D134 in the apo trajectory of ecRNH compared to trajectories containing an Mg^{2+} ion in either the A or the B site. Hydrogen bonds were considered formed if the donor-acceptor distance was less than 3.1Å and the donor-hydrogen-acceptor angle was less than 25°.

Chapter 6

Long-timescale RNase H simulations

6.1 Introduction

The potential for synergy between molecular dynamics simulations and NMR has been recognized since the earliest simulations were conducted [160; 161; 162]. Although NMR has been a fruitful technology for the study of the internal dynamics of biological macromolecules at atomistic resolution, interpreting this information in terms of structural changes has often proven challenging. To put the problem simply, NMR is highly effective at characterizing the amplitude and frequency of motion, but has many fewer tools for characterizing its direction. Combined MD-NMR studies offer the possibility of bridging this gap via direct simulation of processes observable by NMR; however, this approach has historically been constrained by the short timescales accessible to simulation due to the limitations imposed by computational cost [163].

A variety of computational approaches have been developed in order to improve the ability of simulations to observe processes that occur on timescales that are measurable experimentally and relevant biologically. Progress in this field is reviewed in depth in [264] and will be described only briefly here. Among the most critical sources of progress has been the development of efficient parallel-decomposition schemes allowing simulations to be distributed over many individual CPUs (e.g. [265; 214]), thereby enabling access to much longer timescales than would be feasible on a single machine. A fertile area of research focuses on the development of "extended-sampling" techniques that exploit properties of statistical mechanics to expand the regions of conformational space that can be explored in an atomistic molecular dynamics simulation. Replica exchange [266], accelerated molecular dynamics [267], umbrella sampling [268; 269], metadynamics [270; 271], a variety of "steered" or targeted approaches [272], and very large numbers of short simulations run by distributed computing [273], have been described. However, extended-sampling methods all share the property that they improve the quantity of conformational space sampled at the expense of directly simulating processes of biological interest; the temporal dimension of a traditional unbiased MD trajectory is largely lost. Alternatively, one may retain the temporal dimension by reducing spatial resolution instead; coarse-grained modeling found extensive use in early simulation work, when nonpolar hydrogen atoms were often not explicitly represented (e.g. [274; 275]), and has since been developed for the purpose of modeling large macromolecular complexes [154]. However, many dynamic processes of significant interest are not well modeled by such reduced representations.

If neither spatial nor temporal resolution can be compromised, the remaining alternative is to increase the efficiency with which unbiased MD simulations can be carried out. An intuitively appealing method replaces explicit representation of solvent—typically the largest contributor to the computational cost of each timestep—with an implicit-solvent model intended to reproduce the average properties of bulk solvent [276] (reviewed in [277]). This approach is attractive in eliminating large numbers of calculations of "uninteresting" solvent-solvent interactions, but typically performs poorly in benchmarks comparing the resulting macromolecular dynamics on short timescales to experimental ensembles [278]. Improvements in performance of explicit-solvent simulations derive primarily from improvements in parallelization efficiency and from improvements in hardware performance.

Alternative computer architectures have also been pursued as means of dramatically improving simulation performance and thereby gaining access to long timescales. Several attempts at developing special-purpose hardware for MD simulations were made during the 1990s [279; 280], of which the longest-running is the MD-GRAPE project [281]; however, these projects never achieved broad appeal, likely due to their inaccessibility to the broader community. IBM's BlueGene/L system, while not specifically intended for MD, nevertheless proved itself capable of high performance, while suffering the same difficulties of relative inaccessibility [282]. Currently, two major alternatives exist to simulations on traditional CPUs: a special-purpose computer called Anton, developed by D.E. Shaw Research [283]; and graphical processing units (GPUs), particularly via NVidia's CUDA libraries [284; 285; 286]. Highly efficient computations can be performed on readily available commercial GPU hardware. Anton, which has been shown to be capable of up to 1ms continuous trajectories [42] and up to 4ms of aggregate simulation time [287], has been made available to the academic community. Herein the results of long-timescale simulations of the RNase H family are reported, first of 1μs simulations performed on commodity hardware, and also of simulations of up to 14μs performed on the Anton platform. Although local unfolding of some members of the RNase H family was observed, likely due to force field errors at long timescale in a region of the protein with highly unusual structure, these results nevertheless describe the longest reported trajectory to date of a protein from a thermophilic organism and thus afford unique opportunities for benchmarking the sampled ensemble against NMR data.

Table 6.1: **RMSD of chemical shift predictions compared for 100ns and 1μs trajectories of ecRNH and ttRNH.**

Simulation	Cα	Cβ	C'	HN	Hα	N
ttRNH X-ray	1.00	1.14	1.21	0.44	0.32	2.70
ttRNH 100ns	0.68	1.01	1.06	0.35	0.23	2.17
ttRNH 1μs	0.70	1.01	1.06	0.35	0.25	1.99
ecRNH X-ray	0.74	—	—	0.39	0.25	2.51
ecRNH 100ns	0.70	—	—	0.36	0.25	2.25
ecRNH 1μs	0.75	—	—	0.36	0.27	2.32

RMSD of chemical shift predictions for the two bacterial RNase H homologs for which corresponding NMR data are available. Simulation data is reproduced from [184]. Experimental data for ecRNH is from [135] and for ttRNH is from [72].

6.2 Microsecond-timescale simulations of ecRNH and ttRNH

Simulations reaching 1μs total simulation time were initiated for ecRNH and ttRNH under conditions equivalent to those used for the previously described 100ns trajectories (300K, pH 5.5). Interestingly, these longer-timescale trajectories did not significantly improve agreement with experiment when compared with either backbone amide S^2 order parameters (Figure 6.1) or with predicted chemical shifts (Table 6.1) when compared to the 100ns trajectories.

The difference between preferred handle-region populations in ecRNH and ttRNH is significantly reduced in the 1μs trajectories relative to the 100ns equivalents (Figure 6.2). While the ttRNH open-state population is consistent with that observed at shorter timescale, ecRNH experiences many more open-to-closed transitions in the longer trajectory and is most likely converging toward equal populations. This observation is in contrast to the 100ns data, which suggested that higher simulation temperature would be required in order to surmount the energy barrier between the two states and therefore equalize populations.

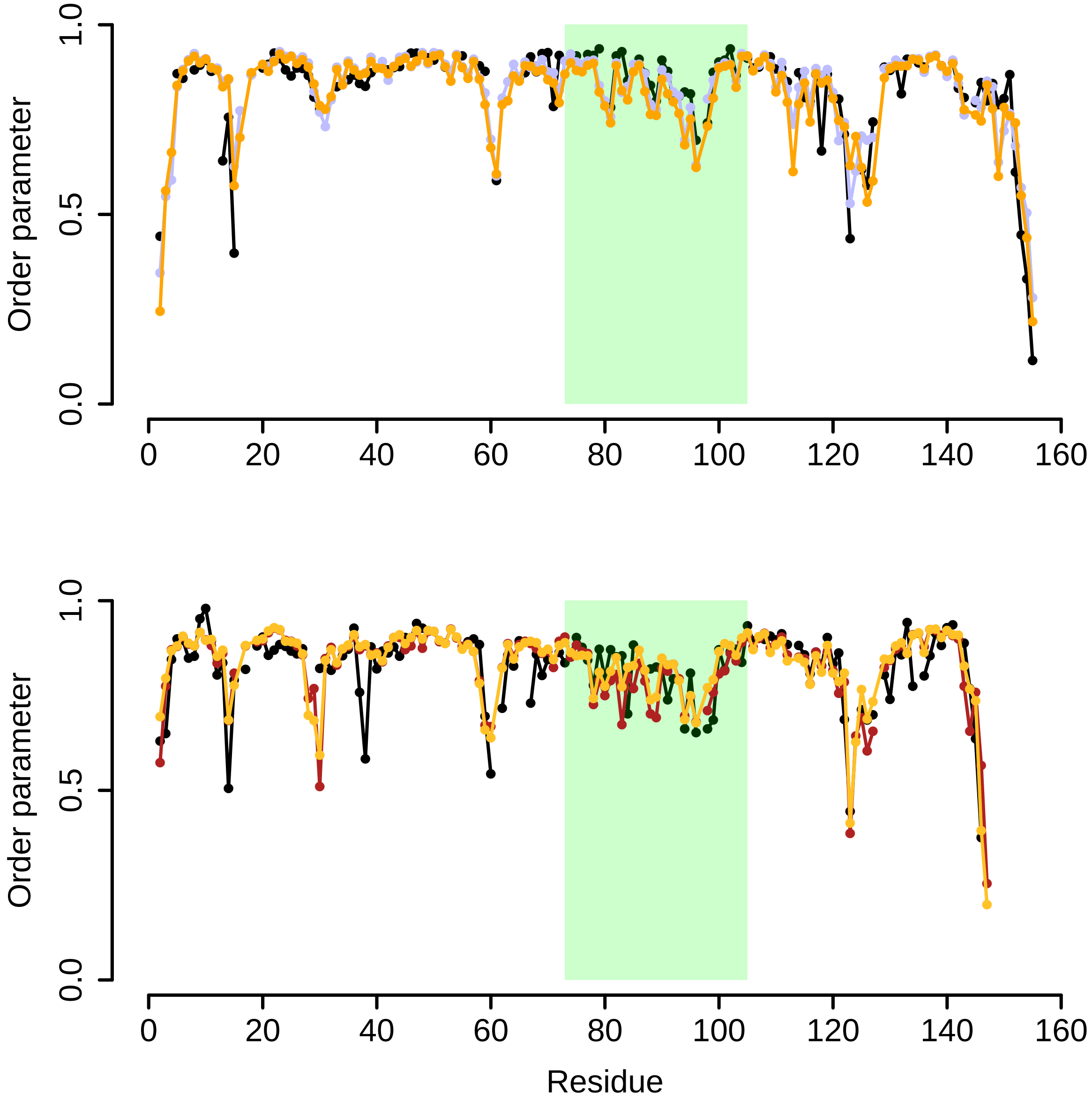

Figure 6.1: **Comparison of experimental and simulated S^2 in 100ns and 1μs trajectories for ecRNH and ttRNH.** Backbone amide order parameters are shown for ecRNH (top: experiment, black; 100ns, blue; 1μs, orange) and ttRNH (bottom: experiment, black; 100ns, red; 1μs, yellow). The green box indicates helices B, C, and the handle loop. Error bars are omitted for clarity.

However, the basic observation that ecRNH populates the open state more frequently than ttRNH is consistent with the 1μs data.

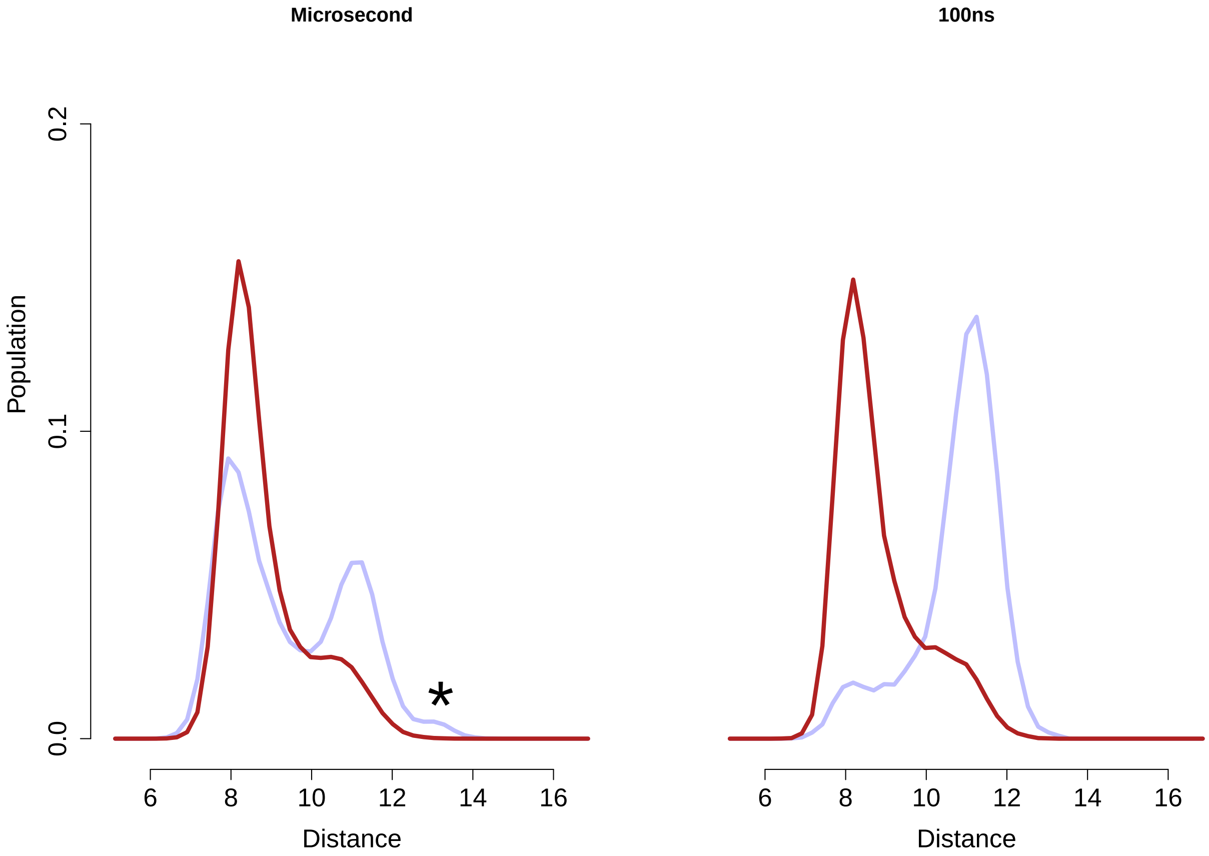

Figure 6.2: **Comparison of handle-region populations in 100ns and 1μs trajectories for ecRNH (blue) and ttRNH (red).** The microsecond-length trajectory (left) has a significantly higher population of the closed state than does the 100ns trajectory (right) for ecRNH, while ttRNH shows little difference. Asterisk indicates the transient population in ecRNH in which the hydrophobic packing of the handle is disrupted.

A brief excursion from the typical open-closed dynamic is observed in the ecRNH trajectory. This corresponds to transient water invasion of the hydrophobic spine of the ecRNH handle loop, in which the SASA of V98 transiently increases as it loses contact with the surface of the W85 ring. Although this excursion persists for only 20 ns in the 1μs trajectory, similar behavior at longer timescale serves as the initiation of unfolding of the handle loop hydrophobic core.

6.3 Simulations of RNases H on the special-purpose hardware platform Anton

6.3.1 System setup

Trajectories were initiated on Anton from the last frames of either the 100ns or 1μs trajectories for ecRNH and ttRNH in AMBER99SB. Trajectories in other force fields were initiated from 10ns NVT simulations run on commodity hardware. Trajectories for all other RNase H homologs were initiated from the last frame of a 100ns NVT run. No significant differences were observed that could be attributable to the length of the trajectory used to generate an initial conformation for Anton runs.

Due to the stability of the ecRNH and ttRNH trajectories at 1μs, trajectories on Anton were not expected to undergo dramatic conformational changes, but rather to provide more thorough sampling of relatively slow motions. Because an extensive dataset had already been collected on the behavior of ecRNH and ttRNH and their point mutants in the AMBER99SB force field, Anton trajectories were first initiated in this force field despite recent evidence that recent modifications, particularly in sidechain dihedral potential surfaces, provide improved agreement with NMR benchmarks [149; 288; 186].

6.3.2 Instability of the handle region at long timescale in the presence of R88

The handle loop of ecRNH (as well as its close relative soRNH) was found to be unstable at long timescale in Anton trajectories. Local unfolding consistently occurred at approximately the same total simulation time of 2.5-3.5μs. Although local unfolding influences the behavior of the C-terminal tail, which is positioned close to the handle in the crystal structure, the rest of the ecRNH protein other than the handle and helix C does not exhibit unusual instability. The RMSD of the non-handle regions of the protein is correlated to that of the handle, but never exceeds 3Å (Figure 6.3). Relevant dynamics for a representative unfolding trajectory (initiated from the 1μs commodity-hardware run) are shown in Figure 6.4.

The mutant ecRNH V98A was observed to unfold more quickly than ecRNH WT, typically within 1μs. For this reason, it was used as a test platform for variations in simulation conditions in an attempt to identify conditions under which the handle loop was relatively stable. However, changes in force field (AMBER99SB-ILDN, CHARMM22*, CHARMM36), water model (TIP4P), electrostatics parameters, integrator, and ensemble did not significantly alter the unfolding behavior observed in ecRNH.

6.3.3 Unusual structural features of the handle loop

The structure of the handle loop is highly unusual, as was commented upon by both of the groups who independently solved the ecRNH structure that represented, at the time, a novel fold. Of the eleven residues that make up the handle loop—G89 to N100 in ecRNH—four occupy the left-handed helical region of Ramachandran space, and only one of those four is a glycine. It is known that substitution of the left-handed K95 with glycine, the residue that occurs in this position in ttRNH, increases the stability of ecRNH by 1.9 kcal/mol,

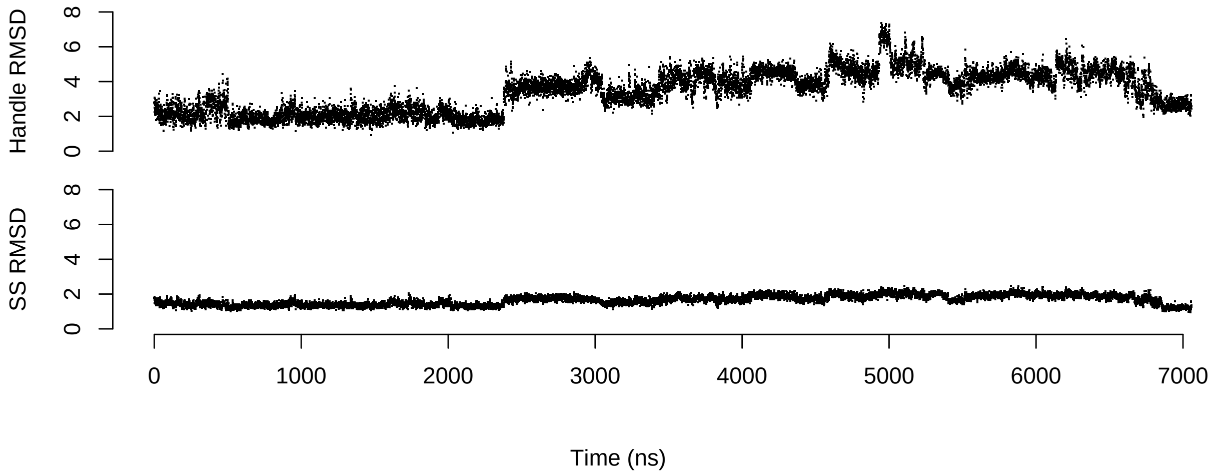

Figure 6.3: **Time courses of RMSD in the handle and core from a representative ecRNH trajectory in which handle unfolding occurs.** RMSD values compared to the crystal structure (2RN2) for an Anton trajectory of ecRNH initiated from a prior run of 1μs. Unfolding occurs at approximately 2.5μs. (Top) RMSD of helices B, C, the handle loop, and the first turn of helix D (residues 73-104), which encompasses the aromatic cluster that defines the handle-region hydrophobic spine. (Bottom) RMSD of all atoms in secondary structure.

the single largest stability increase from a point mutation identified during the course of extensive study of reciprocal ecRNH/ttRNH mutants [98]. The left-handed dihedral for residue 95 is required in order to maintain the native orientation of the backbone in the handle loop; its loss in the ecRNH trajectory occurs simultaneously with overall handle unfolding (Figure 6.4). The left-handed conformation of W90 is also unusual [289], though well-conserved in those RNase H homologs with a tryptophan or other aromatic residue at this site that have been characterized crystallographically.

After approximately 7μs of simulation time, the ecRNH handle loop re-forms a compact structure with relatively low Cα RMSD, but which lacks both the K95 and W90 left-handed dihedrals (Figure 6.3 and 6.5). In this conformation W85 packs directly against V101, ejecting V98 from the hydrophobic spine characteristic of non-retroviral RNase H handle regions.

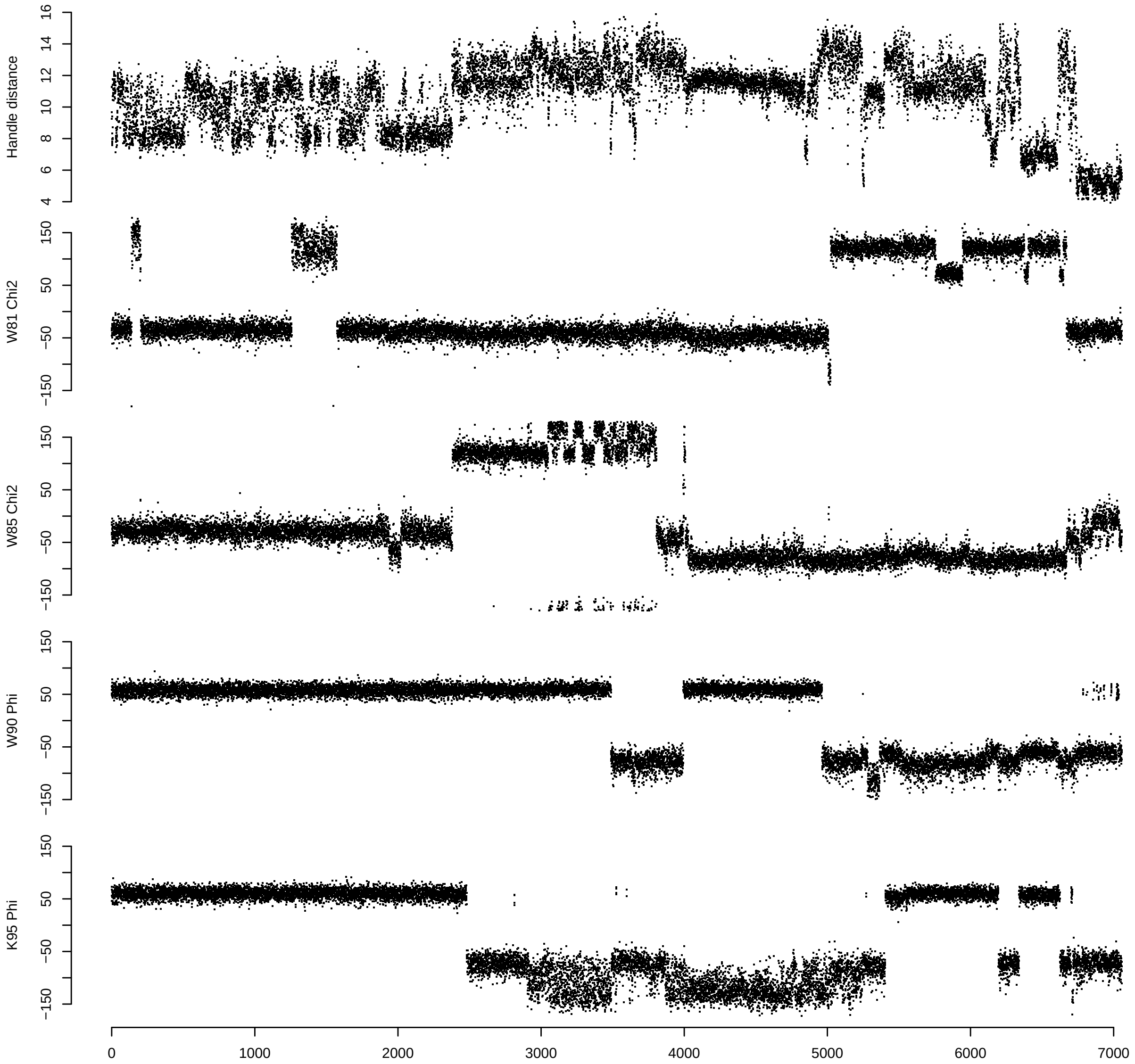

Figure 6.4: **Time courses of structural features in a representative ecRNH trajectory in which handle unfolding occurs.** Structural features that report on handle-loop hydrophobic packing are shown for the Anton trajectory of ecRNH initiated from a prior run of 1μs. Unfolding occurs at approximately 2.5μs and is nearly simultaneous with the loss of the positive ϕ-angle (left-handed helix in Ramachandran space) for K95.

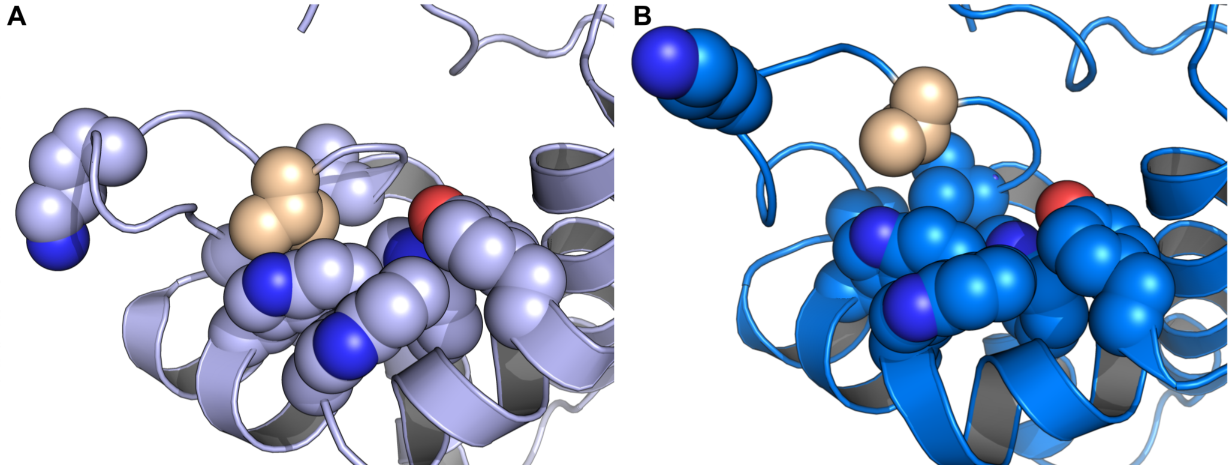

Figure 6.5: **Hydrophobic packing in the ecRNH handle region.** (A) The native hydrophobic packing in ecRNH; V98 is shown in brown. (B) The relatively compact, low-RMSD alternate form sampled in the Anton trajectory at approximately 7μs. In this conformation V98 has been ejected from the hydrophobic core, W85 directly packs against V101, and the characteristic left-handed dihedrals for W90 and K95 have been abandoned.

6.3.4 Handle-loop conformational plasticity in remote RNase H homologs

The simplest explanation for the observed dynamics in the RNase H handle loop is force field error; however, the possibility that transient unfolding is a real event that is simply oversampled in the Anton trajectories cannot be excluded. Although the structure of the handle loop is very well reproduced among a large number of structures of bacterial homologs, as well as of the hsRNH-substrate complex, RNase H domains from retroviruses exhibit a much larger degree of conformational plasticity. Crystallographic characterization of retroviral RNase H domains that natively contain helix C and the handle loop is a long-standing challenge, with multiple examples solved only upon deletion of this region [290; 210]. Recently several structures of full-length RNase H homologs from retroviruses have become available and reveal a degree of conformational variability in the handle loop that far exceeds what has been observed previously in the family [213; 212; 211].

Interestingly, simulations of these retroviral RNases H on the 100ns timescale reveal conformations that somewhat resemble those explored in the "unfolded" state of ecRNH. Additionally, an "atypical" metagenome-derived RNase H structure has also recently been solved, and the hydrophobic packing observed in this structure resembles that sampled in unfolded ecRNH [238]. A characteristic feature of this alternative packing arrangement is the contact formed between W85 and the residue at position 101 (valine in ecRNH, isoleucine in retroviral homologs), with the residue at position 98 ejected from the core and exposed to solvent (Figure 6.6). Because ttRNH lacks a beta-branched residue at position 101, it is unable to assume the same unfolded conformation; instead, its unfolding process initiates with the loss of the left-handed dihedral of W90. In addition, trajectories of ecRNH V101R/Q105E did not unfold within 3μs, reinforcing the importance of V101 in the handle-region hydrophobic cluster.

Prior work on Anton has identified similar large-scale disruptions in hydrophobic packing, known as "cracking", in the EGFR kinase protein [291]. In that case a reorganization was observed into an alternate conformation that was also known *a priori* from crystallography, while the RNase H homologs have no such known variation in handle conformation. However, the trajectories in the EGFR case were longer than those calculated in the present work, so it is possible that a similar reorganization into a relatively stable alternate conformation would be observed given additional sampling time.

6.3.5 Stability of the handle region at long timescale in the presence of N88

Unfolding events were observed in all three examples of RNases H containing arginine at position 88, which was previously identified as a critical determinant of handle-region dynamics at the ps-ns timescale. However, the two RNase H homologs that contain asparagine

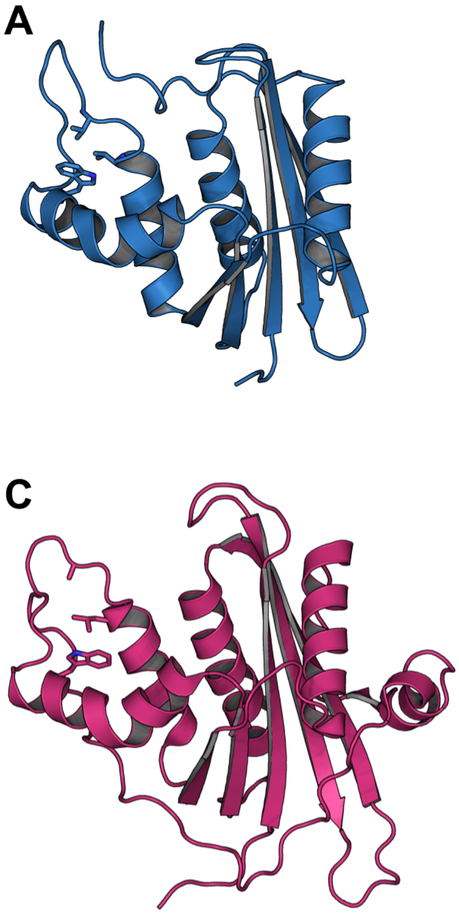

Figure 6.6: **Handle-region hydrophobic packing in remote RNase H homologs and unfolded ecRNH structures.** (A) and (B) Two representative conformations from the unfolded handle loop from the ecRNH Anton simulation. (C) Structure of the "atypical" RNase H homolog recently identified by metagenomics, of unknown source organism [238]. (D) Representative conformation from the 100ns simulation of an RNase H domain from the prototype foamy virus [213]. Different rotamers are observed for W85, but the hydrophobic packing is similar.

at position 88 and therefore form stable hydrogen bonds between the N88 sidechain and the backbone of the residue at position 91 both retained stable, native hydrophobic packing at long timescales, possibly due to the resulting stabilization of the left-handed dihedral conformation of W90. The ctRNH protein was stable for 14μs total simulation time; this trajectory thus represents the longest reported simulation of a protein from a thermophilic organism. The hsRNH protein was simulated for 5.5μs and also remains folded, although it experiences slightly larger fluctuations compared to ctRNH. No large-scale conformational changes were observed in the handle region for either protein (Figure 6.7).

As expected from inspection of the RMSD values, the handle-distance distributions reflect population of a single state, slightly broader in hsRNH (Figure 6.8).

Interestingly, the ctRNH trajectory almost exclusively populates the *gauche-* conformation for the χ_2 angle of W81 (Figure 6.9), previously identified in short-timescale simulations as a potentially binding-incompetent conformation. This behavior is consistent with patterns observed from mutants at short timescale: in the presence of both a beta-branched residue at position 101 and a glycine insertion, the *gauche-* conformation will be favored (*vide infra*, 4.4.3). However, steric considerations suggest that the *gauche-* conformation should clash with substrate, making it a surprising observation that a wild-type protein should favor this conformation so strongly. Notably, the ctRNH handle loop was not resolved in the crystal structure for this protein and was therefore modeled using ttRNH as a template (with W81 in the *trans* conformation); however, if the maintenance of this conformation is an artifact of errors in the starting structure, it is an extraordinarily persistent one. The *gauche-* conformation is also sampled in the hsRNH trajectory, although with lower frequency. The ring-current effects from this highly persistent rotamer should be sufficient to induce chemical shift changes of a magnitude large enough to exceed the error of current prediction methods; it remains to be seen whether this conformation in ctRNH can be validated experimentally.

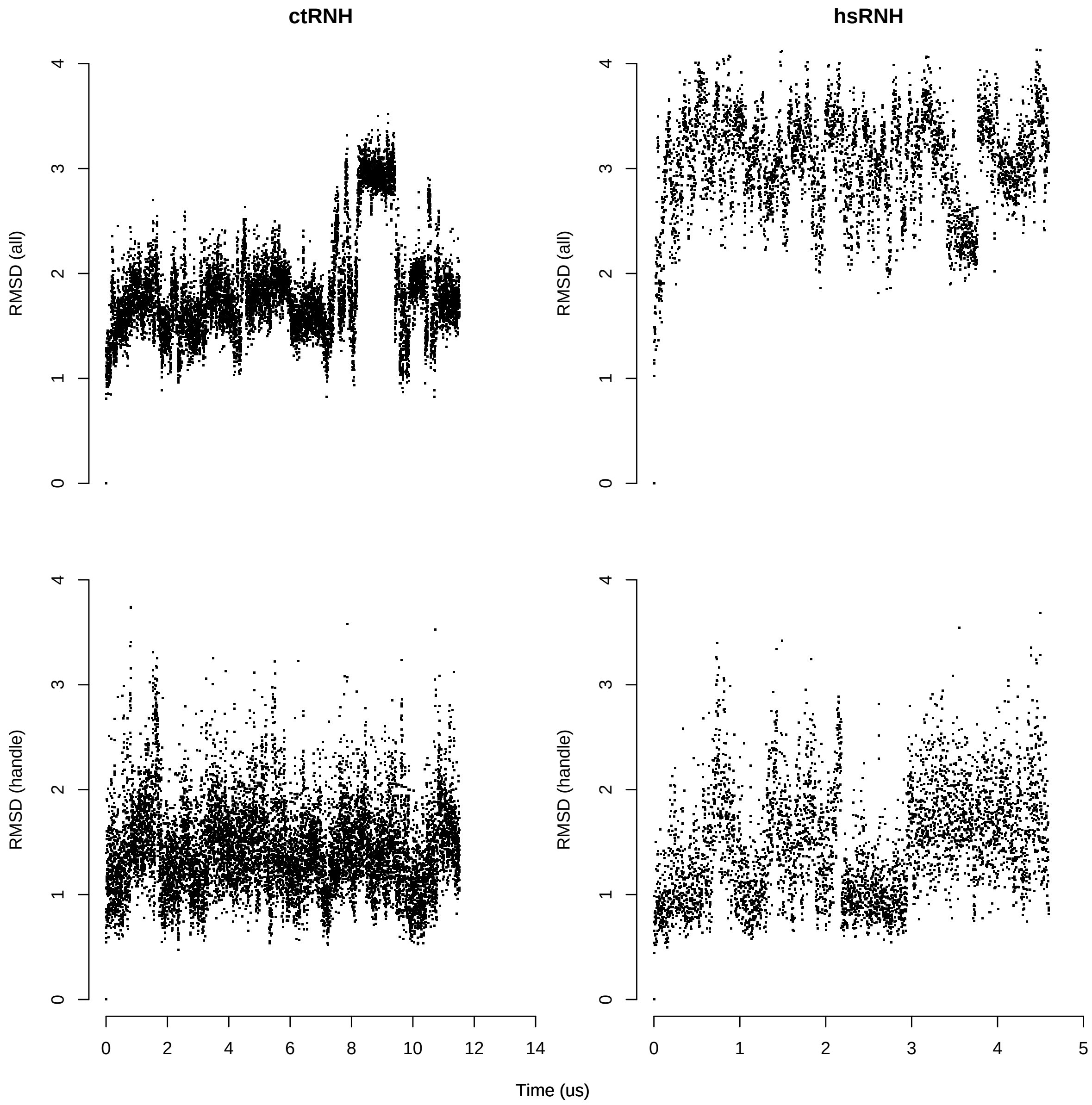

Figure 6.7: **Time-dependent RMSD for Anton simulations of N88-containing proteins.** C_α RMSD values to the starting structures are shown for ctRNH (left) and hsRNH (right), for the whole protein (top) and for helices B and C and the handle loop (bottom) (residues 73-104). No large changes are observed for the handle loop of either protein. The excursion for ctRNH at approximately 8μs represents transient melting of the the C-terminal end of helix E.

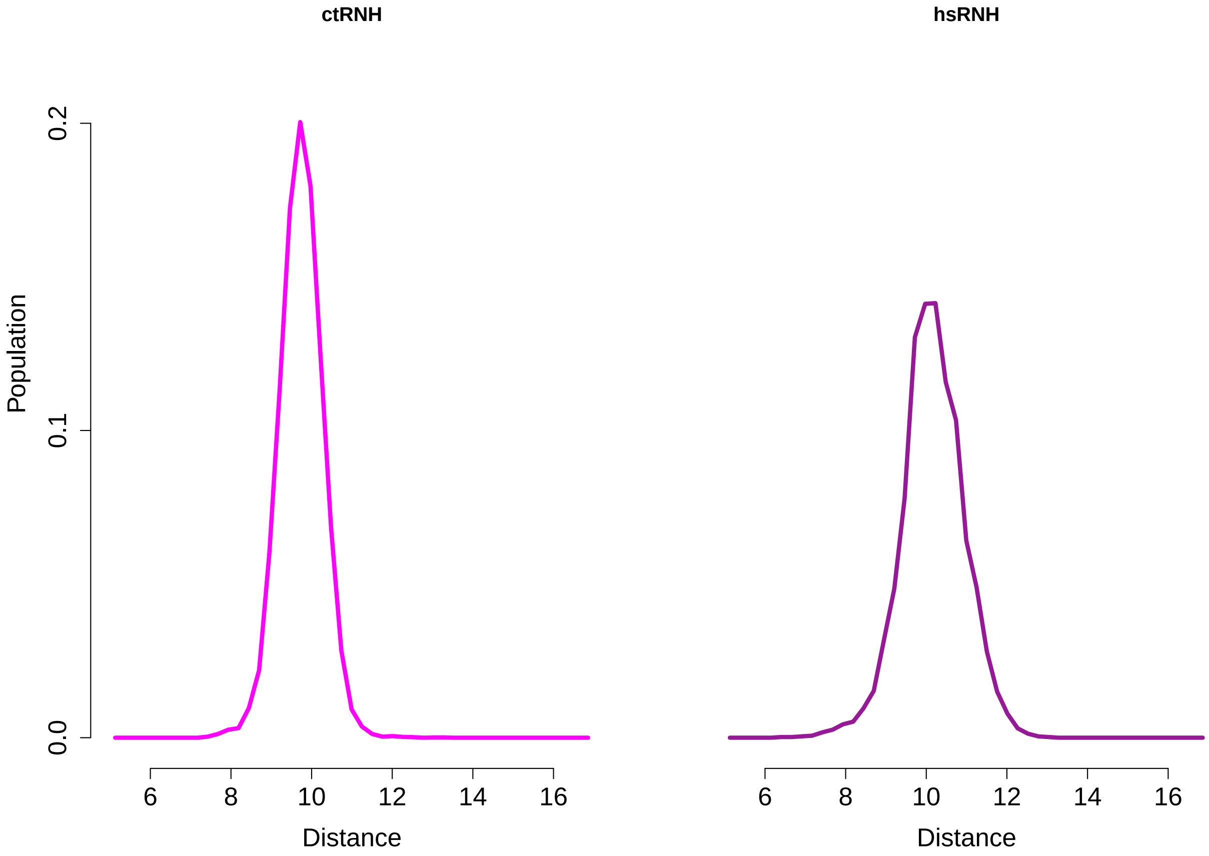

Figure 6.8: **Handle-distance distributions for Anton simulations of N88-containing proteins.** The distance distributions are calculated for ctRNH (left) and hsRNH (right). The ctRNH distribution is noticeably narrower, consistent with its lower time-dependent RMSD in this region.

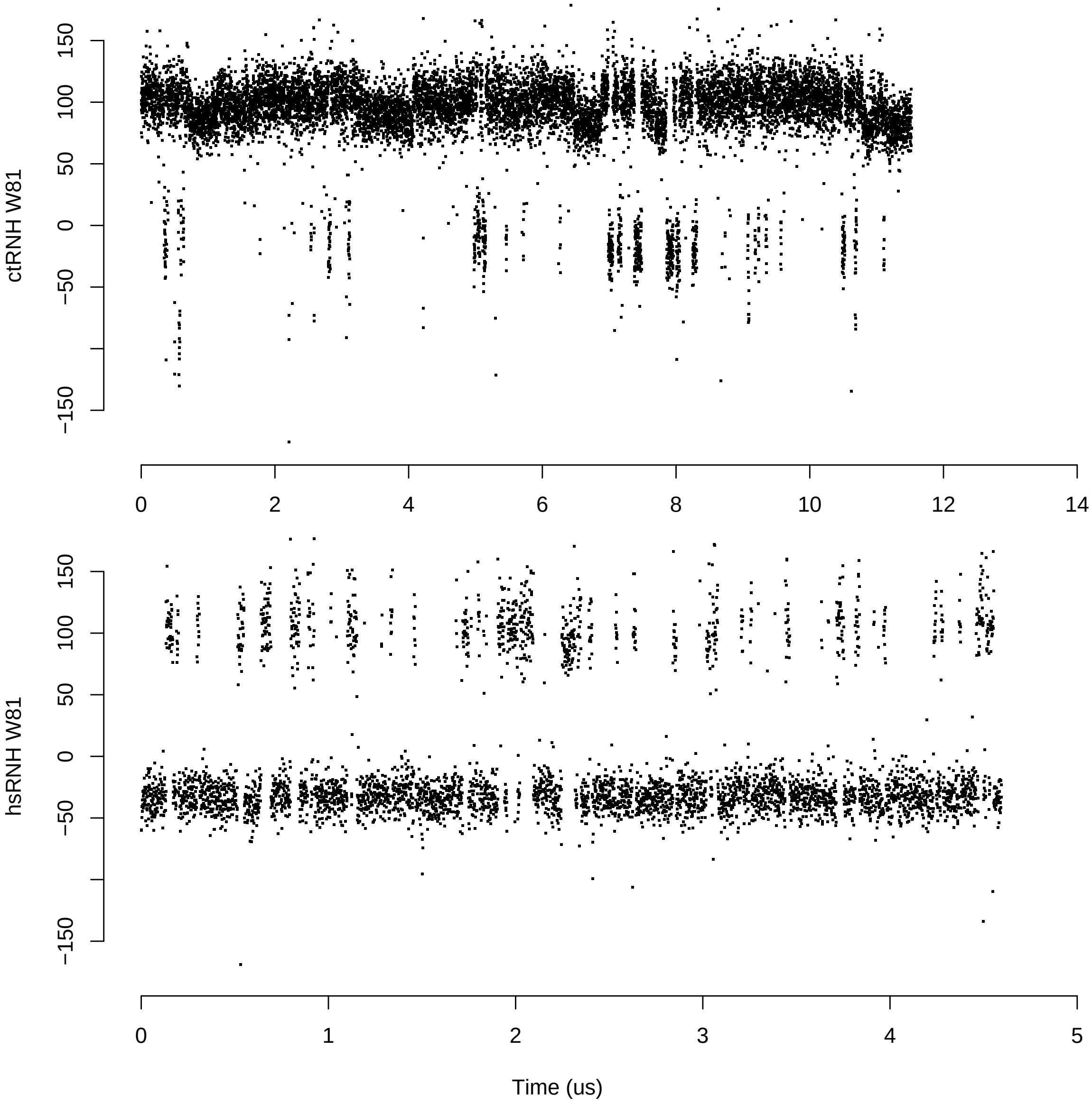

Figure 6.9: **W81 dynamics for Anton simulations of N88-containing proteins.** The W81 χ_2 angle as a function of time is shown for ctRNH (top) and hsRNH (bottom).

Chapter 7

NMR spectroscopy of ctRNH

7.1 Introduction

Extensive prior NMR investigations have been performed on ecRNH and ttRNH, motivated by the hypothesis that differences in dynamics might explain the otherwise somewhat mysterious differences in catalytic activity between these two structurally very similar proteins [131; 132; 133; 134; 135; 72; 73]. Furthermore, the structural bases of their different thermostabilities have been investigated in detail and attributed largely to then-unexpected differences in the ΔC_p values of the two proteins [30]. The reduced ΔC_p observed in ttRNH has been attributed to residual structure in the unfolded state [114]. Chimeric proteins in which the folding core of ttRNH is exchanged with the exterior of ecRNH and vice versa suggest that the origin of this thermostabilization lies in the hydrophobic core [113]. Furthermore, the residual structure can be selectively disrupted by charged mutations at specific sites that form loci of residual structure [114].

Interestingly, ctRNH—derived from a moderately thermophilic organism with a preferred growth temperature of around 40°C—retains the low ΔC_p value of ttRNH but has a T_m almost identical to that of ecRNH [101]; its residual structure can be probed by mutation and overlaps with but is distinct from that of ttRNH [99].

Due to its position as an "intermediate" in thermodynamic behavior between ecRNH and ttRNH, ctRNH was not anticipated to have dynamic behavior that differed significantly from either of its better-studied homologs. However, we have shown that simulations on both short (*vide infra*, Chapter 3) and long (*vide infra*, Chapter 7) timescale reveal distinctive dynamic behaviors in this protein that may reflect an alternative evolutionary solution to the problem of thermal adaptation of conformational dynamics relevant to substrate binding. This distinctive behavior is in simulation entirely determined by a single residue at position 88 in the sequence, conserved as arginine in ecRNH and ttRNH but occupied by an asparagine in ctRNH, resulting in rigidification of the handle loop due to the formation of stable hydrogen-bonding interactions between the asparagine sidechain and the backbone of a neighboring residue. The position of the mutation in the structure is shown in Figure 7.1.

Note that throughout the preceding chapters, the residue numbering in all RNases H has been determined for consistency with ecRNH; for this reason, the glycine insertion is G80b rather than G81. Due to software limitations in representing peak annotations, the present chapter will use numerical numbering for ctRNH consistent with that present in PDB ID 3H08. The mutation site is located, in this numbering scheme, at position N89.

The ctRNH protein thus represents an appealing target for further study by NMR methods. Herein we begin the process of characterizing the dynamic behavior of wild-type and N89R ctRNH by NMR and comparing the results to those previously obtained for ecRNH and ttRNH. This study represents a unique opportunity to validate a "blind" prediction from simulation of the dynamic behavior of a novel mutant. Furthermore, because a 14μs trajectory of the wild-type ctRNH has been produced using the Anton platform, this study will facilitate comparison between long-timescale simulation and experimental study of thermal adaptation.

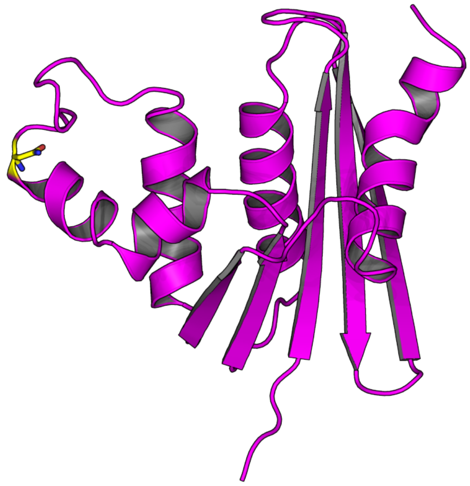

Figure 7.1: **Location of the N89R point mutation in ctRNH.** Isotopically labeled samples were produced and NMR spectra collected for the wild-type ctRNH, which has asparagine at position 89, and for its N89R mutant. The location of the mutation at the C-terminus of helix C and immediately preceding the handle loop is shown here in yellow on the last frame of a 100ns simulation trajectory (used to illustrate the preferred N89 rotamer, which was not the orientation found in the crystal structure.)

7.2 HSQC spectra

Sensitivity-enhanced $^1H -^{15} N$ heteronuclear single-quantum coherence (HSQC) experiments were conducted on both the WT and N89R ctRNH proteins at 0.95mM and 0.73mM respectively at 299K. Excellent chemical shift dispersion and high signal-to-noise ratios were obtained for both proteins, suggesting that both are well-folded globular structures and that little if any degradation occurred. The spectra are highly superimposable, revealing relatively minimal shift perturbations. Importantly, dispersion is equivalent in the wild-type and the mutant, suggesting that the mutation is well-tolerated and does not disrupt or substantially destabilize the structure.

Although HSQC spectra are primarily used to study backbone amide groups, signals corresponding to sidechains that contain N-H groups also appear and can have properties of

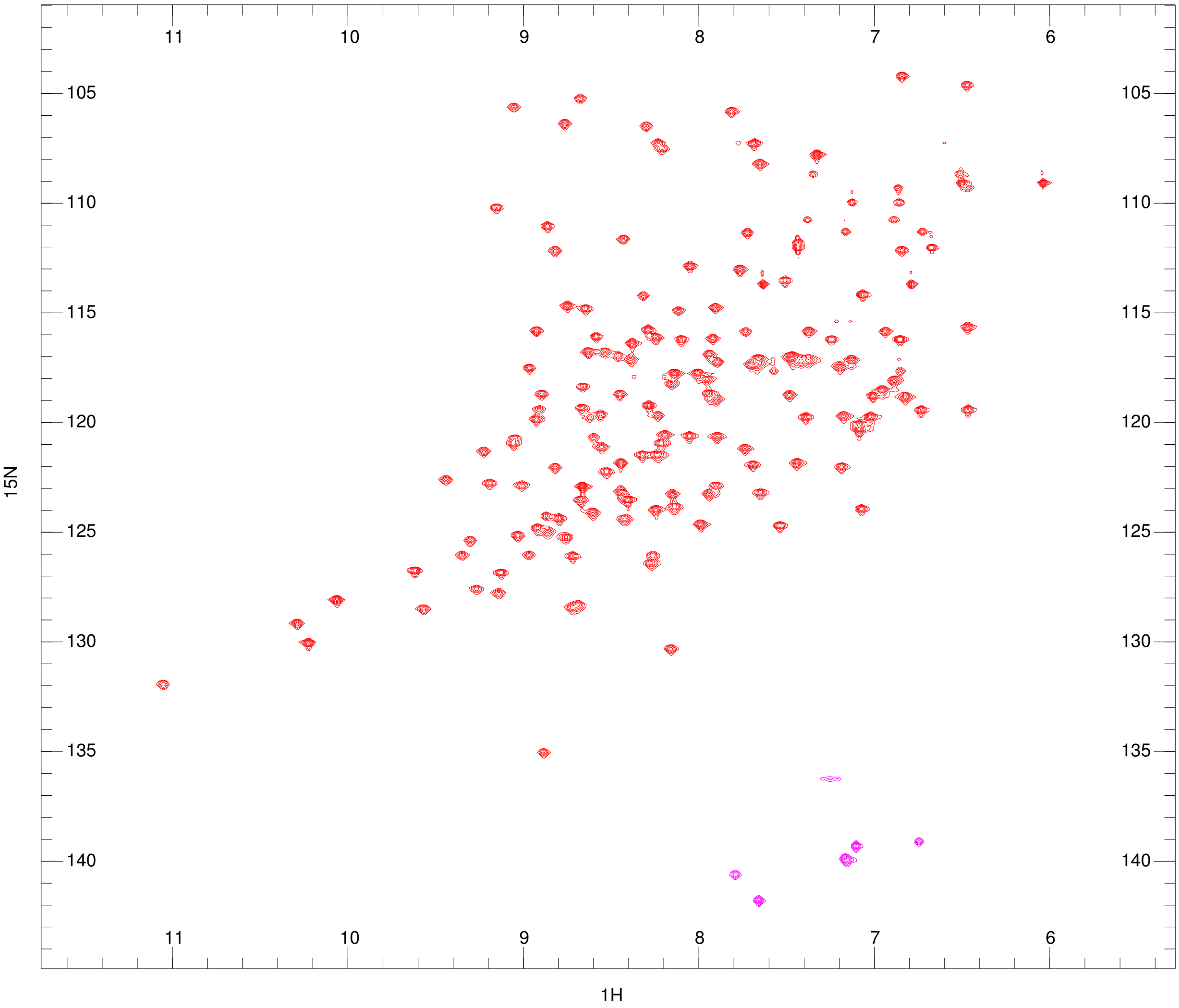

Figure 7.2: **HSQC spectrum of wild-type ctRNH.**
Excellent chemical shift dispersion is observed in the spectrum and the peak count is consistent with expectations based on the protein sequence. Arginine residues are shown here as negative, folded peaks (pink) due to limited spectral width.

interest. Distinctive regions of the spectrum are typically occupied by signals corresponding to the indole amides of tryptophan rings, the guanidinium groups of arginine sidechains, and the amide groups of asparagine and glutamine sidechains. Locations of the tryptophan and arginine sidechains, whose N^{ϵ} sidechain groups are visible in the spectrum, are shown in Figure 7.4.

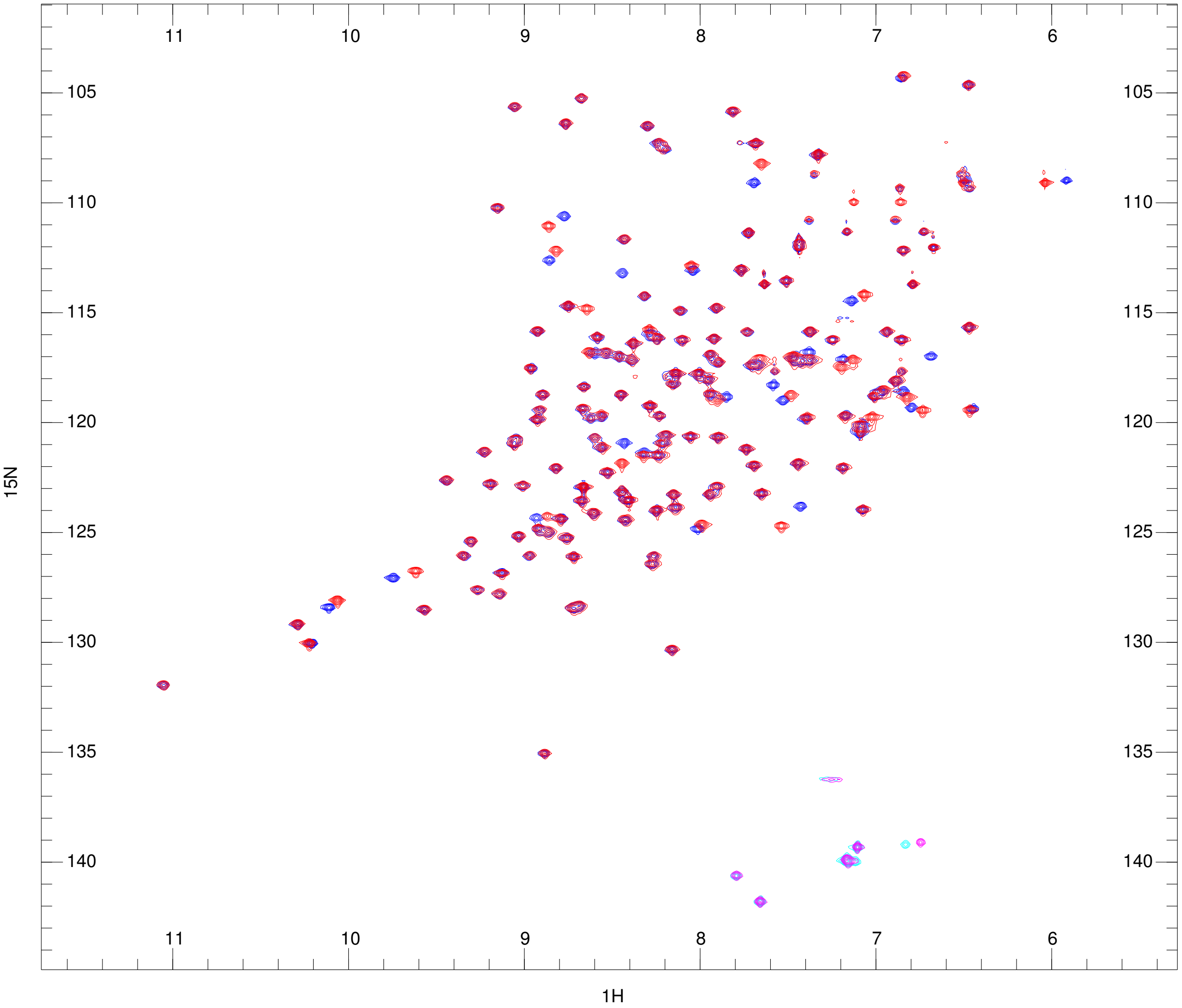

Figure 7.3: **Superposition of HSQC spectra of wild-type and N89R ctRNH.** The mutant spectrum (blue, cyan) superposes well on the WT spectrum (red, pink), indicating that the mutation is well-accommodated and nondisruptive.

Five tryptophan N^{ϵ} signals were observed in each spectrum, of which two display perturbation due to the mutation, which is located in a cluster of four tryptophans that form part of the hydrophobic spine in the handle.

Signals from the arginine $N^{\epsilon} - H^{\epsilon}$ bonds in the sidechain guanidinium groups were also observed (shown here as negative and folded, due to limited spectral width). Dispersion

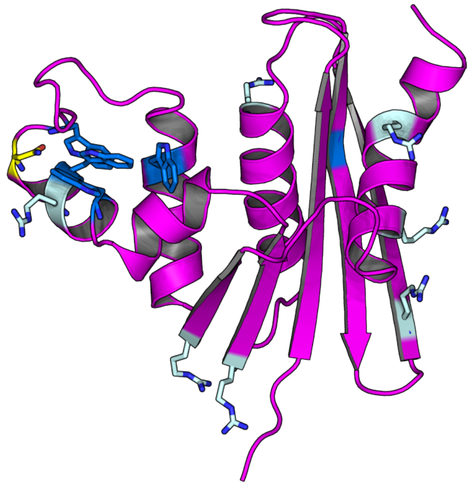

Figure 7.4: **Location of tryptophan and arginine sidechains in ctRNH.** Locations of tryptophan and arginine sidechains, whose $N^{\epsilon} - H^{\epsilon}$ bonds are visible in distinctive regions of the HSQC spectrum.

in the arginine region is slightly worse than what has been observed in ecRNH [243], with multiple overlapped or unusually shaped peaks, including the one attributed to the presence of the N89R mutation.

Because the HSQC experiments were performed several weeks before the completion of the backbone assignment experiments, additional spectra were collected after these experiments to confirm the integrity of the samples. No changes in spectra were observed up to four months after preparation of the samples.

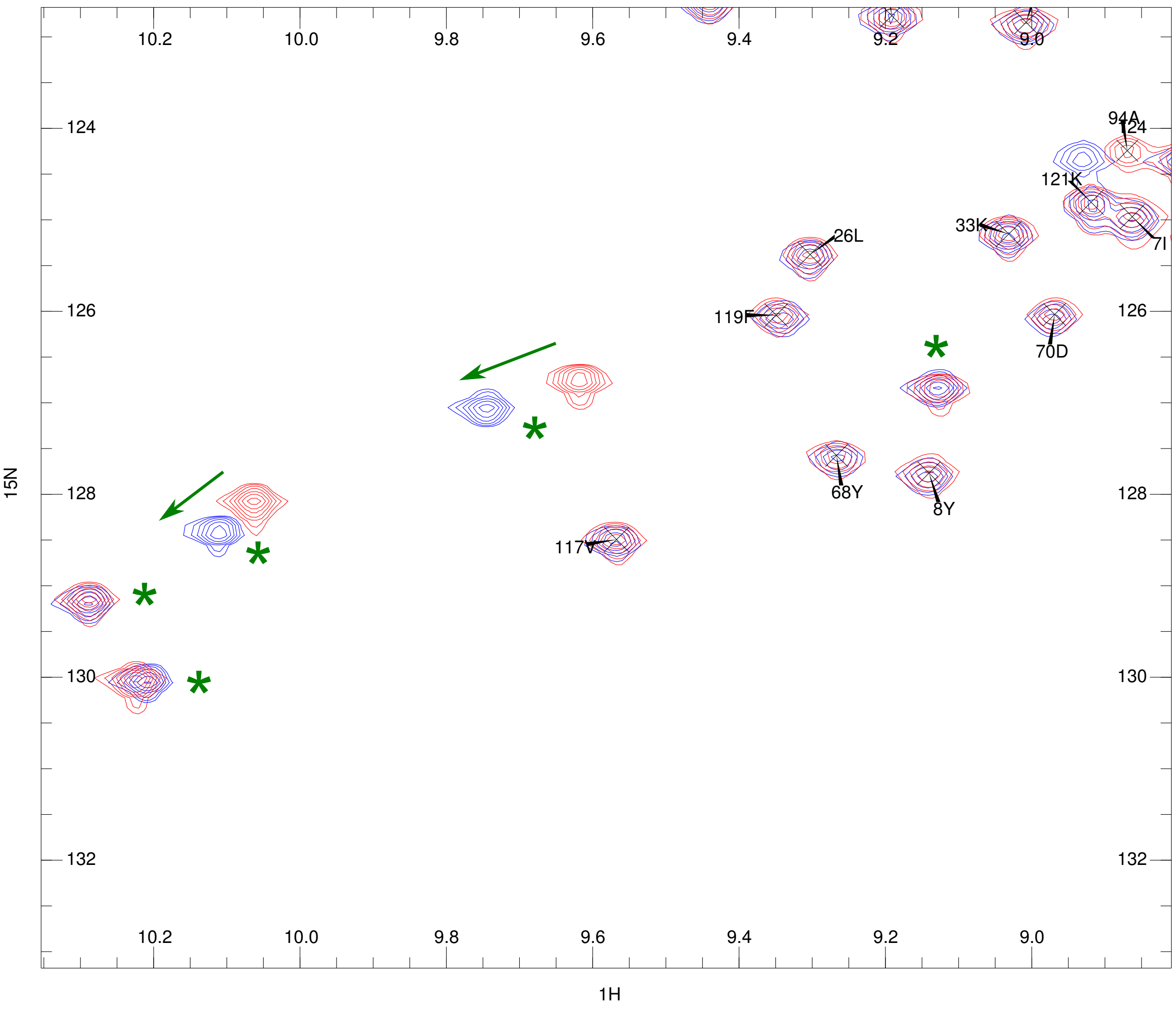

Figure 7.5: **The characteristic tryptophan region of the ctRNH HSQC spectra.** Resonances determined to be associated with the amide group of the tryptophan indole ring in the wild-type (red) and N89R mutant (blue) are labeled with green asterisks. Significant perturbations are observed for two of the five tryptophan residues in the protein, of which four are located in the hydrophobic cluster that forms the handle loop.

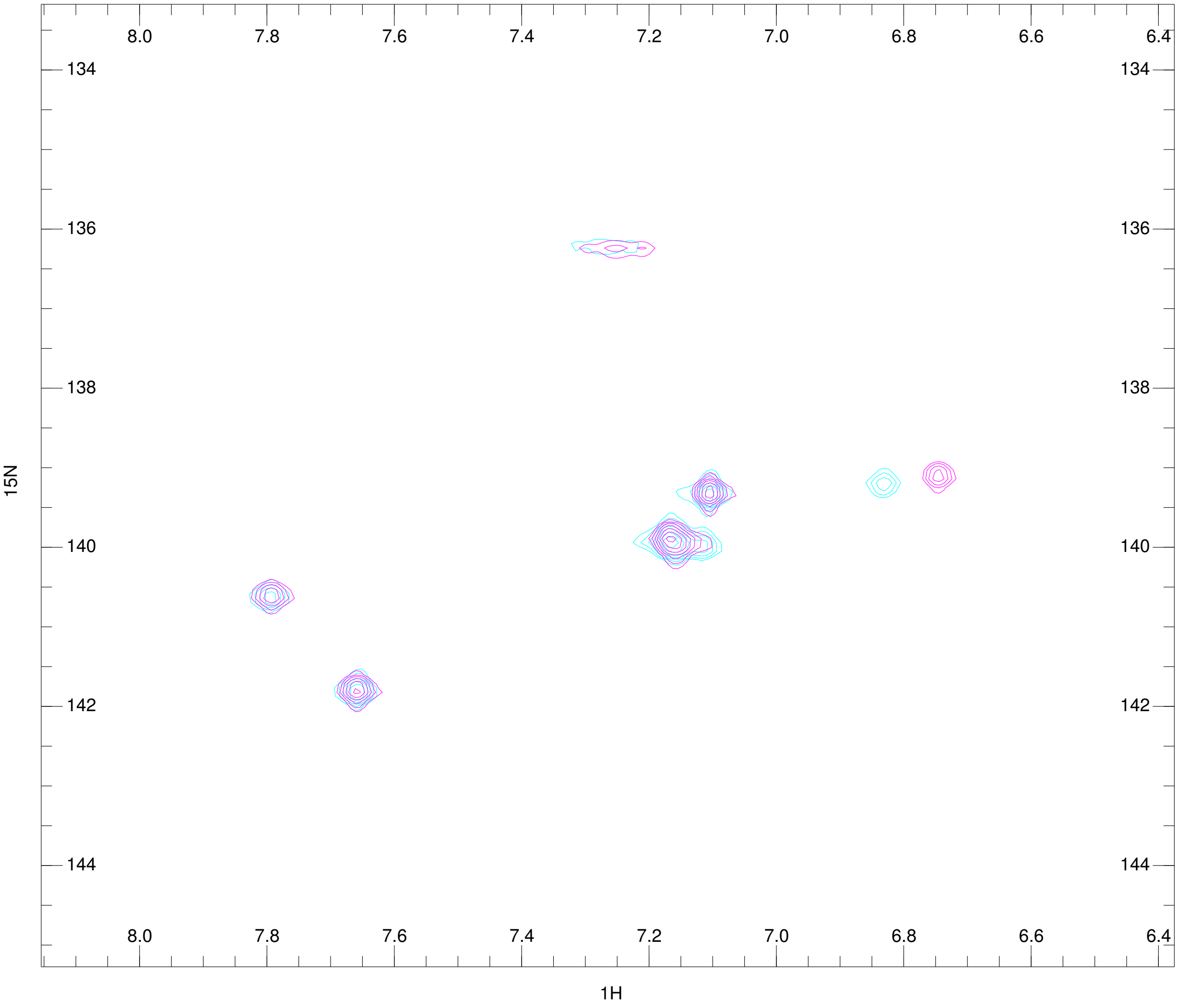

Figure 7.6: **The characteristic arginine region of the ctRNH HSQC spectra.** The arginine region is shown for the wild-type protein (pink) and the N89R mutant (cyan). The arginine residues are not as well-resolved as they are for ecRNH, resulting in peak overlap that obscures the signal associated with the mutation.

Table 7.1: **Unassigned non-proline residues in ctRNH WT.**

Location	Residues
N-terminus	G0
αA-β4 loop	K60 (tentative), E61
αC	W86 (tentative), V87
αD	E107
β5	H120, V122
αE	L140
C-terminus	S146

Locations of non-proline residues that remain unassigned in ctRNH WT. These residues localize mainly to the termini, to the αA-β4 loop, and to the αC helix immediately preceding the mutation site. E61, V87, and E107 could not be definitively assigned due to peak overlap, and K60 and W86 could be assigned only tentatively. Note that G0 is a non-native residue that is produced as an artifact of the purification procedure and is not expected to have a significant effect on the behavior of the protein.

7.3 Backbone assignments of handle-region residues

NMR spectra collectively sufficient to complete backbone assignments of both proteins have been collected at 299K. HSQC, HNCACB, and HNCACOCB spectra were collected for both proteins and are sufficient for assignments of the large majority of residues. Approximately 95% of the non-proline backbone resonances have been assigned in wild-type ctRNH (Figure 7.7). Seven non-proline residues remain unassigned; their identities are summarized in Table 7.1.

Interestingly, one region presenting assignment difficulty localizes to the αA-β4 loop. Relaxation in this loop, centered around K60, has been observed in ecRNH [72], suggesting that dynamics may explain the difficulty in assigning these remaining residues. Importantly, the entire handle loop (excluding P98) could be assigned, although some residues in αC remain elusive.

Sites that exhibit distinctive chemical shifts in the WT and N89R proteins localize primarily to the handle region and to αD. Importantly, only a small perturbation was

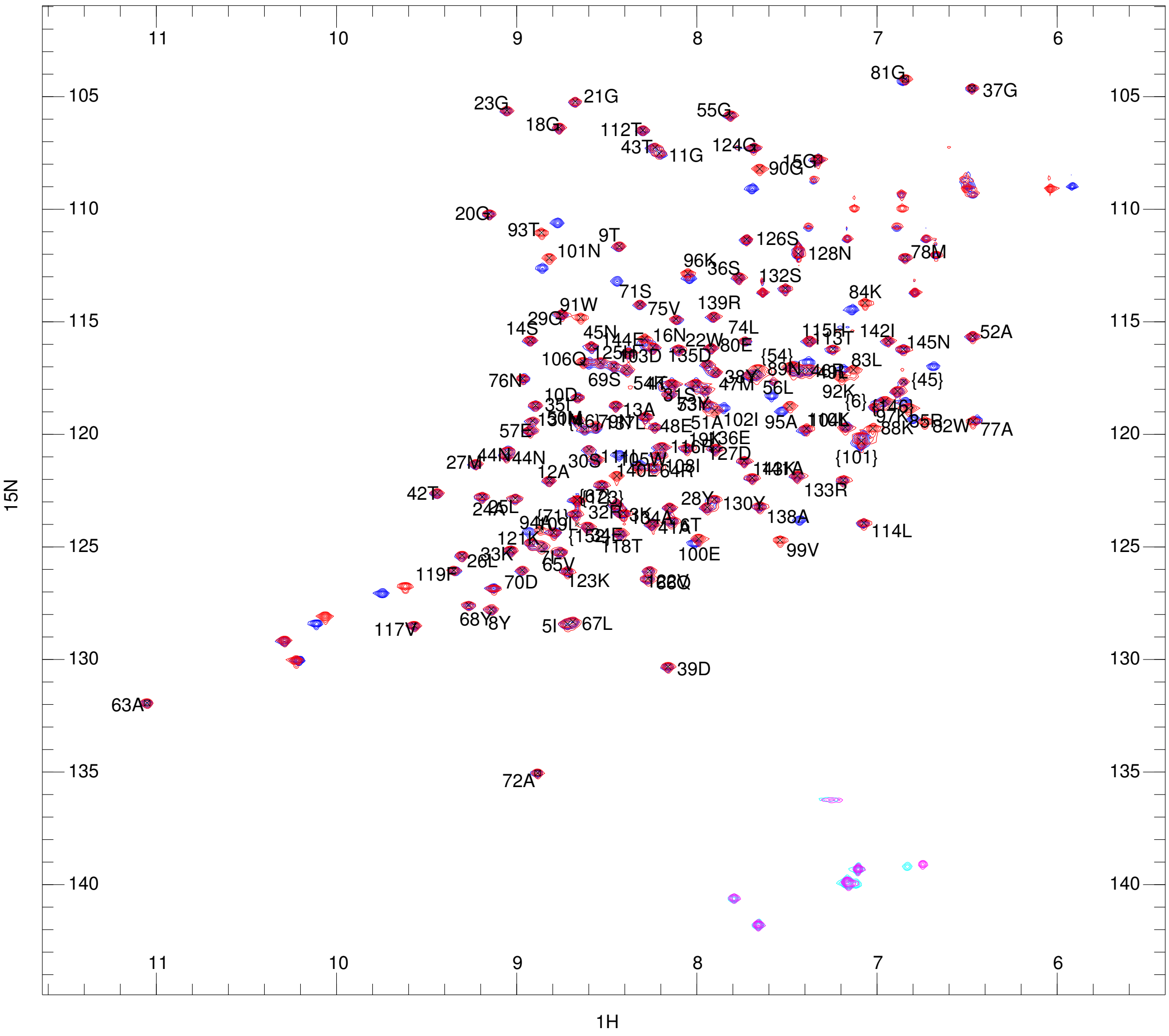

Figure 7.7: **Backbone assignments for ctRNH WT.**
Backbone assignments are shown for ctRNH WT (red) superposed on the spectrum of ctRNH N89R (blue). Resonances associated with sidechains—arginine, tryptophan, asparagine, and glutamine—are not assigned.

observed for the glycine insertion G81, suggesting that the local environment for this site was not significantly altered. Figures 7.8, 7.9, and 7.10 show shift perturbations in three distinct regions of the spectrum.

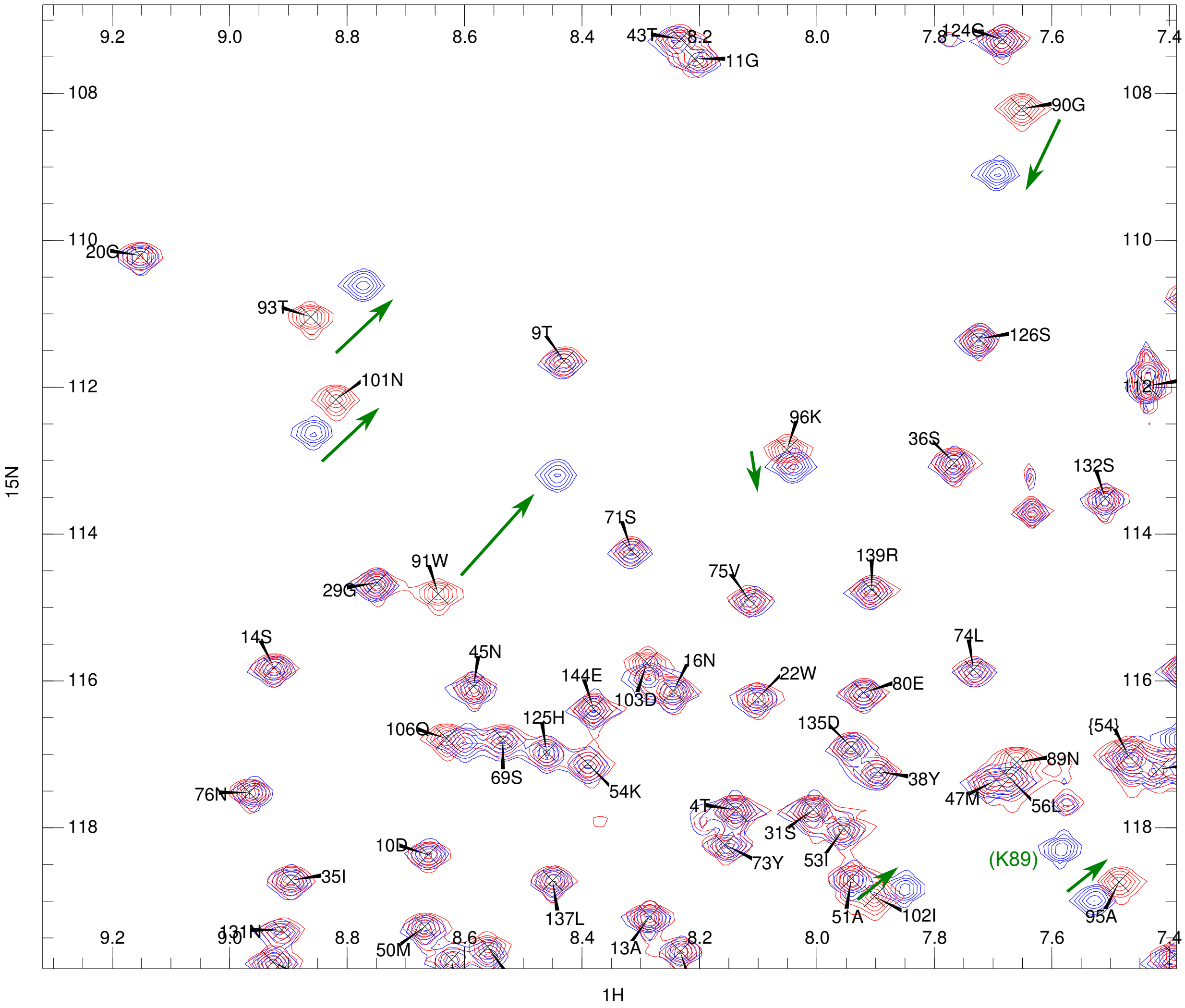

Figure 7.8: **Shift perturbations in ctRNH WT and N89R, part 1.** Signals for ctRNH WT (red) and N89R (blue) are shown with green arrows connecting corresponding resonances for those cases in which shift perturbations are observed.

7.4 Comparison to dynamically averaged chemical shift predictions

Dynamically averaged chemical shift predictions for ctRNH WT and N89R were produced using Sparta+ based on the Anton trajectories of each protein and were used for comparison

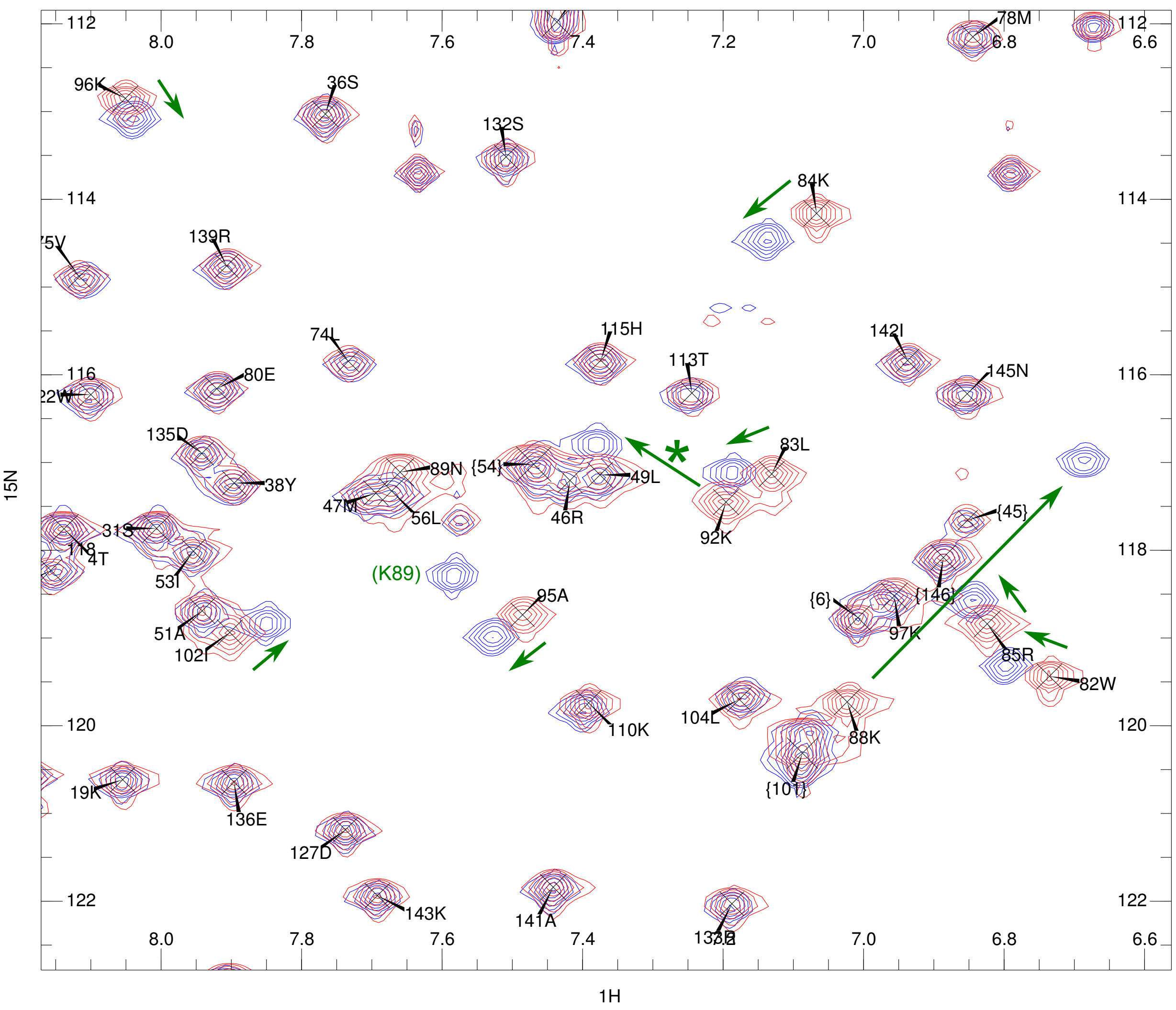

Figure 7.9: **Shift perturbations in ctRNH WT and N89R, part 2**
Spectra are colored as in Figure 7.8. The green asterisk marks the "diagnostic" K92 shift.

of predicted and experimentally identified perturbed residues. In the case of N89R, in which the handle loop unfolds on an approximately 2μs timescale in a manner consistent with that observed in ecRNH, predictions were averaged over the 2μs prior to unfolding. By contrast, ctRNH WT did not unfold on the timescale sampled, permitting averaging over the full 14μs trajectory. (Results in the following discussion would not be significantly affected if sampling were limited to 2μs.) RMSD values for the amide proton and nitrogen shifts are

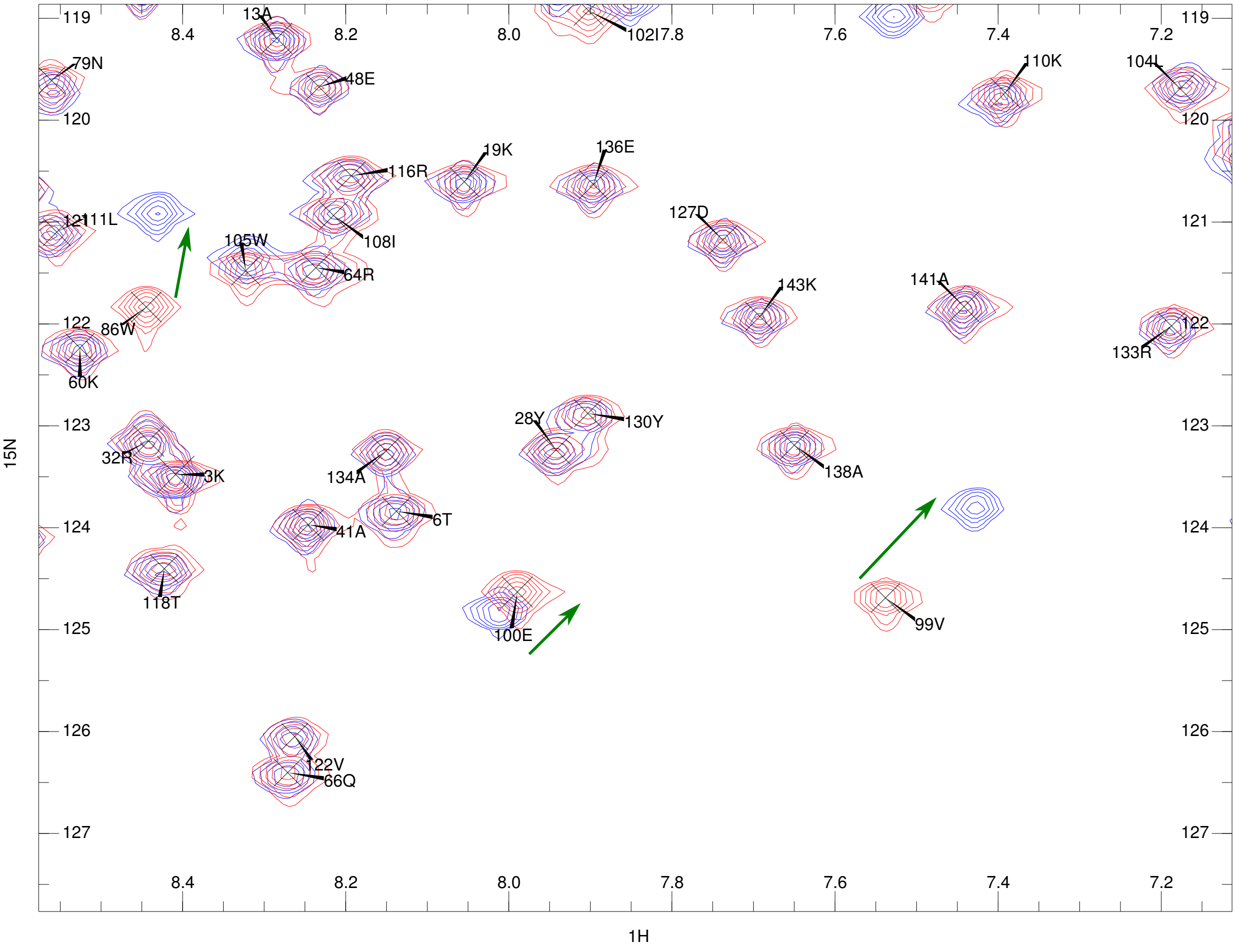

Figure 7.10: **Shift perturbations in ctRNH WT and N89R, part 3**
Spectra are colored as in Figure 7.8.

shown in Table 7.2; for nitrogens, these RMSD values are comparable to those observed in predictions for ecRNH and ttRNH made from shorter-timescale simulations, while the ctRNH predictions for protons are slightly worse (Table 3.1). Values for ctRNH WT and N88R were extremely similar (though both sets of experimental shifts were referenced to very similar structures, into which the missing residues of the handle loop in PDB ID 3H08 had been rebuilt).

Qualitatively speaking, the simulations correctly predict that shift perturbations due to the introduction of the N89R mutation are primarily localized to the handle loop and

Table 7.2: **Backbone amide and nitrogen chemical shift prediction RMSDs for ctRNH WT and N89R.**

Protein	HN RMSD	N RMSD
ctRNH WT	0.48	2.37
ctRNH N88R	0.48	2.36

RMSDs for the backbone amide proton and nitrogen chemical shift predictions from long-timescale Anton simulations, compared to experimental measurements for ctRNH WT and N89R.

its immediate environment. However, although perturbations are also predicted in the region of β5 and at the N-terminus of αC, such differences are not observed experimentally. This observation suggests that differences in conformational sampling between the two trajectories introduce noise in attempts to predict shift perturbations even at very long simulation timescales. Predicted and experimental shift perturbations are shown on the ctRNH structure in Figure 7.11.

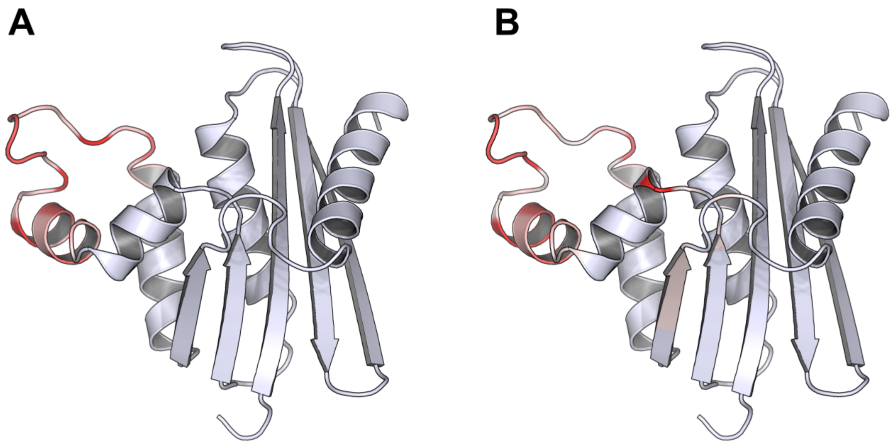

Figure 7.11: **Structural view of N-H shift perturbations in ctRNH WT and N89R.** (A) Shift perturbations derived from experimental data. (B) Shift perturbations derived from dynamically averaged predictions based on Anton trajectories of 14μs (WT) and 2μs (N89R). Red indicates increasing perturbation in either the proton or nitrogen dimension. Sites with no change are shown in light blue.

In simulation, the backbone amide of K92 forms a stable hydrogen bond to the sidechain of the native N88, but interacts only with solvent in the N89R mutant, affording an opportunity to identify a diagnostic chemical shift perturbation to validate this structural variation between the two proteins. Signals associated with the K92 backbone amide in the HSQC spectra are illustrated in Figure 7.9, revealing a perturbation of 0.2 ppm downfield in the proton dimension, whereas the predicted chemical shift perturbation is 0.7 ppm in the opposite direction. Notably, it is the wild-type prediction that shows an error of almost 1ppm; the N89R prediction is quite close to the experimental value. The expected predictor error for amide protons is approximately 0.5 ppm [184]. Overestimation of the magnitude of chemical shift perturbations in shift-prediction datasets is consistent with other examples of RNase H mutants studied to date (P. Robustelli, unpublished data); however, the discrepancy in sign remains a matter of interest in understanding the behavior of these proteins. In particular, the observed backbone chemical shift of K92 suggests that the simulated population of the hydrogen-bonded conformation may be too large.

The largest chemical shift perturbations observed experimentally occur at sites W90 and K88, which are immediate neighbors of K92. This observation suggests that the conformation of the handle-loop hinge has indeed been altered; however, it is unclear given the present data whether the conformation has changed in ways not sampled by the simulations, or the predictions simply lack sufficient resolution to pinpoint single-site changes induced by weakly perturbing mutations. It must be emphasized that the handle-loop hinge is in a highly unusual conformation that features both G90 and W91 in left-handed regions of Ramachandran space. It is possible that this unusual conformation is inadequately modeled by force fields (see Section 6.3.3). Since chemical shift predictors rely on both physical models and on machine-learning approaches based on empirical datasets, it is likely that they too are weaker in modeling rare conformations. Notably, predictions for the K92 site are poor in the ecRNH and ttRNH cases as well.

These results emphasize the need for a holistic approach to understanding the predicted effects of mutations on conformational dynamics, rather than identifying specific, diagnostic shift perturbations expected to report on the presence or absence of predicted conformational states.

Chapter 8

Conclusions and future directions

8.1 Conclusions

8.1.1 Summary

The present work demonstrates the productivity that can be achieved through close coupling between experimental observations by nuclear magnetic resonance spectroscopy and molecular dynamics simulations in the study of evolutionary adaptation of functionally relevant conformational dynamics.

In the first study presented herein, prior NMR work on the handle region, a functionally important substrate-binding loop, in ecRNH and ttRNH was reanalyzed in the context of an interpretive framework developed through comparative analysis of MD simulations of all five RNases H from cellular organisms. This analysis yields a model of the conserved dynamic mode that is both more intuitive than the previous analysis and more consistent with prior work on the correspondence of preferred conformational states with optimal temperatures for protein homologs adapted to differing thermal environments. Furthermore, an additional, previously unsuspected alternative handle-region dynamic mode was identified in RNase H homologs that have recently been structurally characterized, and a single residue was shown to be sufficient for determining the dynamic behavior of this loop through the formation of diagnostic hydrogen-bonding interactions. The picture that emerges from this work is of a loop that can either swing on hinges whose stiffness is adapted to the thermal fluctuations in a protein's native environment, or be buttressed by sidechain-backbone hydrogen bonds. Collectively, these results suggest that, despite high sequence homology among the RNase H proteins studied here, the protein fold permits multiple possible adaptive pathways to balance the competing constraints represented by conformational dynamics and thermostabilization.

In the second study, prior work identifying mutations that increase the activity of ttRNH at ambient temperature was rationalized in the context of differences in sidechain dynamics and preferred rotameric states observed between wild-type ecRNH and ttRNH. Importantly, these effects are distinct from effects on handle-loop conformation, as none of the activating-mutant trajectories result in increased population of the presumptively binding-competent state of the handle loop relative to wild-type ttRNH. simulations suggest that two out of the three activating mutations exert their effects through changes in rotamer preferences of relatively well-packed hydrophobic sites, an adaptive mechanism not widely exploited in the context of protein design.

The third study focused on the dynamics of the carboxylate-containing ecRNH active site as characterized by both experiment and simulation. The simulated dynamics of the carbonyl-containing sidechains were largely validated by comparison to those inferred from recent NMR experiments quantifying motion of these residues at the ps-ns timescale. The active-site residues were found to be rigid in the ps-ns timescale while undergoing substantial conformational exchange upon Mg^{2+} binding, possibly indicating a state of electrostatic preorganization for binding the first metal ion, coupled to dynamic reorganization at longer timescales. This model was supported by simulations of ecRNH in complex with Mg^{2+} and by simulations of other homologs, even very remote in sequence, which collectively suggest that preorganization may be an inherent property of the overall RNase H fold. This work illustrates the advantage of combined MD-NMR studies for understanding the dynamic prerequisites for enzymatic catalysis.

The fourth study investigated the utility of extremely long-timescale molecular dynamics simulations using the recently developed special-purpose hardware platform Anton. Unexpectedly, local unfolding in the handle region—a loop of highly unusual structure—of several RNase H homologs was observed in a variety of force fields and simulation conditions. Interestingly, the two homologs with "single-state" handle-region dynamics determined by a shared hydrogen-bonding pattern in the handle "hinge" did not experience instability, so long-timescale trajectories of the hsRNH and ctRNH homologs were obtained. The ctRNH trajectory, at 14μs, represents the longest reported trajectory of a protein from a thermophilic organism. Surprisingly, ctRNH populates sidechain rotamers suspected based on the second study to be binding-incompetent. This trajectory affords a unique opportunity for comparison to experimental data, as it was performed "blind" from a partial homology model and will be compared in future work to NMR data on this protein.

Finally, the fifth study constituted the experimental characterization of the ctRNH protein, both of the wild-type (cysteine-free) form and of a mutant designed on the basis of the first study to specifically perturb handle-region dynamics in a way that gives rise to diagnostic chemical-shift changes. This is part of ongoing work that will incorporate chemical-shift validation of the 14μs Anton trajectory.

Taken together, the studies presented in this work collectively define a strategy for productive integration of NMR observations with MD simulations, both in the form of a series of relatively short trajectories related by point mutations from a common background and in the form of single continuous long-timescale trajectories.

8.1.2 Binding kinetics of RNases H

The model developed here for the interactions of the RNase H handle region with substrate has important and directly testable implications for variations in binding kinetics within the family. Two conserved dynamic modes were identified in the handle region, determined by the identity of a single residue at position 88 at the C-terminus of helix C: when this site is Arg or Lys, a two-state equilibrium between open and closed states is observed, while an Asn at this site stabilizes a single state roughly intermediate between the extremes defined by the open and closed states. The handle loop has previously been suggested to move as a rigid body in ecRNH and ttRNH; these results suggest that it can either swing on loose hinges, or be buttressed by the sidechain-backbone hydrogen bonds for which an Asn residue at this site is uniquely well-suited. The significance of this residue substitution, and potentially of the corresponding dynamic behavior, is supported by sequence analysis of the RNase H family made possible by modern genomic sequencing methods; in a dataset consisting of nearly a thousand sequence examples, the hydrogen-bond-forming Asn at position 88 is enriched in those sequences annotated as having a thermophilic source organism.

Several studies have demonstrated the close relationship between dynamics observed in an enzyme's apo state and those observed in substrate complexes [292; 293] . Differences in the conformational dynamics of the apo states of homologous proteins could therefore contribute to differences in the kinetics of substrate binding or product release. The binding kinetics of the two classes of RNase H homologs identified here, differentiated by the residue at position 88, are predicted to differ significantly (Figure 8.1). The kinetic scheme for two-state proteins is a two-step process: a conformational selection step in which the substrate binds preferentially to the open state is followed by an induced fit process in which the open handle loop rearranges to form hydrogen-bonding interactions with the DNA strand of the substrate (Figure 8.1A). Because the RNase H protein must discriminate not only

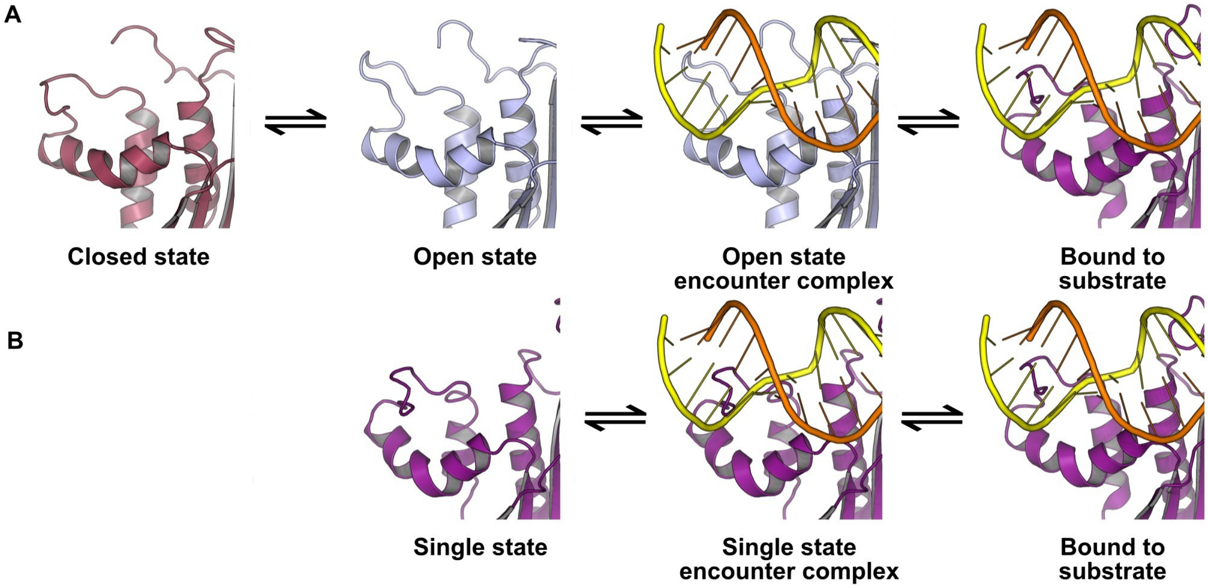

Figure 8.1: **Summary of kinetic schemes for substrate binding in one- and two-state RNases H.**
(A) The kinetic scheme for the interaction of substrate with a two-state handle region, where the open state is the binding-competent state. (B) The kinetic scheme for a single-state handle region, in which the loop is held in a single conformation well-positioned for substrate interactions.

between different types of nucleic acids, but also between the two strands of its hybrid substrate, a two-step process in which an encounter complex quickly dissociates if the strands are misaligned could provide significant regulatory advantage. Altering the relative population of the open state through mutation at sites not directly involved in the substrate-binding interface offers a means for fine-tuning conformational preferences to match both the functional context and the thermal environment. By contrast, the kinetic scheme for the single-state, Asn-containing proteins is a single-step process, as the loop conformation stabilized by Asn-backbone hydrogen bonds is already oriented for productive interactions with substrate (Figure 8.1B).

8.1.3 Relationship between conformational dynamics and electrostatic active-site preorganization

The hypothesis that conformational dynamics are directly related to enzymatic catalysis is an extremely controversial subject in modern enzymology [45; 46; 47]. While much of this controversy centers on the matter of dynamical effects on the chemical step of the catalytic cycle, questions remain regarding the role of dynamics in processes such as binding and orienting substrate and cofactors. High local rigidity is often associated with enzyme active sites [294] and rigid active sites associated with high specificity of binding [295]. Conversely, catalytic promiscuity is associated with flexible active sites, up to the extreme case exemplified by a molten-globule active site [296].

In the present work, lines of evidence from both experiment and simulation suggest that the active site of ecRNH is rigid on the ps-ns timescale, a matter of significance in elucidating the mechanisms through which RNases H interact with catalytically required divalent cations as well as with the substrate. Simulations of diverse RNase H family members, including those found as subdomains of retroviral reverse transcriptases, suggest that this active-site rigidity is not merely a feature of ecRNH in particular, but rather is a general feature of the family and is imposed by the nature of the RNase H fold, which orients the four active-site carboxyl groups in close proximity to one another despite the energetic penalty associated with overcoming the resulting electrostatic repulsion [297]. Furthermore, experiments on the μs-ms timescale suggest the presence of active-site dynamics induced or potentiated by the binding of Mg^{2+}, lending support to the “mobile metal ion” hypothesis of ion-protein interactions in ecRNH [124]. However, simulations in the presence of magnesium ions suggest little change in the ps-ns timescale dynamics exhibited by the protein in the presence of metal, suggesting that electrostatic preorganization on short timescales can coexist with dynamics at longer timescales.

In this work several examples are presented in which conformational dynamics of sidechains distal from the active site are identified as having important effects on activity, either through facilitation of substrate binding or through contributions to the maintenance of catalytically competent local organization of residues surrounding the active site. Significantly, in no case studied here do the hypothesized changes affect the dynamics of the active-site residues themselves, reinforcing the notion that a conformationally rigid active site that is preorganized for effective catalysis can nevertheless be dependent on protein dynamics even in a small and relatively globally rigid globular protein.

8.1.4 Implications for thermal adaptation of in other protein families

Extensive prior work has examined the mechanisms by which proteins adapt to varying properties of the bulk environment. Such studies are typically performed in one of two ways: as surveys of genomic-scale data, in which aggregate properties such as increased numbers of charged residues in proteins derived from thermophiles are identified [13], or as case studies of particular protein pairs or protein families, in which specific residues or structural features can be identified as contributing to thermal adaptation. The present work aims to engage with both approaches by using simulations to explore the conformational space accessible to a larger number of family members, both as wild-type sequences and in the presence of mutations designed to perturb specific dynamics.

In several prior "case studies"—for example, the well-known case of adenylate kinase—the dynamics associated with catalytic activity localize to a flexible "lid" that folds over the active site only in the presence of substrate, thereby isolating the catalytic center from exposure to bulk solvent. The lid-opening dynamic is commensurate with overall kinetic rate and therefore the slower process in the thermophilic protein is sufficient to explain its lower catalytic rate [53]. This case neatly fits the long-standing hypothesis of "corresponding states" [59], according to which homologous proteins should have similar degrees of structural flexibility (and occupy similar overall conformational ensembles) in the optimal temperature ranges of their source organisms.

Although the RNase H family has been cited as an example of a case that does not fit the corresponding-states hypothesis [72; 73; 74], the present work reinterprets prior NMR data within a framework consistent with the concept of corresponding states. In particular, although high-temperature simulations do not significantly increase the open-state handle region population for ttRNH and therefore do not meet the strict corresponding-states criterion, the dynamics exhibited by ecRNH and ttRNH are shown by simulation to sample similar regions of conformational space. This observation provides an interpretive model for the reanalysis of other protein families whose homologs appear not to fit a corresponding-states-like pattern and offers new insight into the possibilities of evolutionary adaptation through fine-tuning of conformational dynamics to match challenges posed by the bulk environment.

8.2 Future directions

8.2.1 Chemical shift predictions as tools for backbone assignment

In the present work, dynamically averaged chemical shift predictions were carried out on wild-type and N88R mutant ctRNH trajectories prior to the initiation of experimental studies of these proteins. However, backbone resonance assignments were performed "blind", without direct coupling to the predicted chemical shift values.

The quality of chemical shift predictions has dramatically improved in recent years, particularly due to improved treatment of ring-current effects due to the circulating π electrons from neighboring aromatic groups [175]. Chemical shift predictions have therefore found increasing use as tools for experimental validation of molecular dynamics simulations [182; 183; 184; 185]. Although CCPnmr Analysis [227], a software suite for resonance assignment and NMR data analysis, includes a facility for supplying predicted shift information to the user based on a static protein structure, the use of shift predictions to guide backbone assignments is not yet common or routine. As exemplified in Chapter 7, the resolution of chemical shift predictions is not yet high enough for use in predicting individual shift perturbations corresponding to small local conformational changes induced by mutation. However, continuing improvement in both prediction quality and in computational power for production of dynamics data or other conformational sampling makes the integration of dynamically averaged chemical shifts with backbone assignments of new proteins a logical forward step.

One difficulty that arises in developing such an approach derives from the different sources of error on experimental and predicted shift values. Prediction errors tend to be heteroskedastic, with larger errors occurring for chemical shift values that represent larger deviations from a given residue type's random-coil shift value. Extreme outliers are inevitably underrepresented given the machine-learning techniques typically used to train the predictors and the biases inherent in the databases from which the training data is taken. This property of prediction tools can exacerbate the difficulty in producing reliable predictions for individual sites of interest. However, errors due to undersampling of the conformational space of the protein being predicted—rather than undersampling of chemical shift space across all proteins—can be significantly improved by the introduction of dynamic averaging.

8.2.2 Experimental studies of RNase H activity and dynamics

The models developed in the present work on the basis of simulation data in conjunction with NMR experiments would benefit from additional experimental data, and several experiments immediately suggest themselves as appropriate approaches for work that builds on the current foundation.

First, binding kinetics and activity measurements of a number of the mutants presented in this work would serve as a rigorous and straightforward test of the functional component of the predictions. Although measuring the catalytic rate of the RNase H reaction can be complicated by substrate dynamics, by the natural weak processivity of the enzyme, and by difficulty with developing efficient assays, the ecRNH V98A mutant has been tested and found to exhibit a reduced K_m relative to wild-type, corresponding to a weaker binding affinity as predicted by this protein's increased population of the closed state in the handle region (P. Robustelli, unpublished data). Additional studies of the activity of ctRNH WT and N88R would be particularly useful as complements to the existing simulations and NMR data. A thorough study of the temperature-dependent kinetics of ecRNH and ctRNH would also be beneficial, as the binding affinity of the two-state proteins would be expected to have a stronger temperature dependence than the one-state proteins.

Second, additional NMR data on this protein is of particular interest as a test of the understanding of the handle region developed here; prior work on ttRNH has suggested that the handle region and active-site loop are coupled on the μs-ms timescale [72], that the dynamics in these regions are associated with the presence of the glycine insertion [73], and that the introduction of a glyine into the corresponding position in ecRNH dramatically reduces activity (J.A. Butterwick and P. Robustelli, unpublished data). In ctRNH WT, the glycine insertion is present but two-state handle-region dynamics are absent; it would be illuminating to examine this protein for relaxation behavior in the αB-αC hinge near the inserted glycine, and to compare this behavior to ctRNH N88R.

8.2.3 Implications for protein design

A long-standing goal of the study of naturally occurring thermostable proteins is an understanding of their properties that can be leveraged in designing novel proteins with both high catalytic activity and high thermostability. Most previous attempts at rational design of thermostability into a naturally thermolabile protein have focused on the introduction of rigidifying structural features such as disulfide bonds and salt bridges [22], though such modifications often directly trade off against catalytic activity [58]. An alternative approach to rigidification achieves thermostability by selectively substituting residues with high crystallographic B-factors [23].

Interestingly, recent attempts to design novel proteins that bind to small target molecules have identified insufficient preorganization of the binding site as a source of weak interactions [298], suggesting that similar deficiencies may underlie the relatively weak catalytic rate enhancement of designed enzymes. However, to date protein design projects typically focus on the design of the binding site or active site, rather than on design of larger-scale dynamics intended to facilitate substrate interactions or product release. The present work represents an early look at the designability of conformational dynamics within a given protein fold and the ability of a computational mutagenesis/molecular dynamics pipeline to explore this space. In future work it is anticipated that design of conformational dynamics may eventually allow rational optimization of thermal adaptation in proteins useful for applications such as biotechnological applications and environmental remediation.

Chapter 9

References

Bibliography

[1] Vihinen M (1987) Relationship of protein flexibility to thermostability. Protein Eng 1: 477–480.

[2] Kopitz A, Soppa J, Krejtschi C, Hauser K (2009) Differential stability of TATA box binding proteins from archaea with different optimal growth temperatures. Spectrochimica Acta A 73: 799–804.

[3] Jaenicke R, Bhm G (1998) The stability of proteins in extreme environments. Curr Opin Struct Biol 8: 738–748.

[4] Thompson MJ, Eisenberg D (1999) Transproteomic evidence of a loop-deletion mechanism for enhancing protein thermostability. J Mol Biol 290: 595–604.

[5] Balasco N, Esposito L, Simone AD, Vitagliano L (2013) Role of loops connecting secondary structure elements in the stabilization of proteins isolated from thermophilic organisms. Protein Sci 22: 10161023.

[6] Gromiha MM, Pathak MC, Saraboji K, Ortlund EA, Gaucher EA (2013) Hydrophobic environment is a key factor for the stability of thermophilic proteins. Proteins 81: 715–721.

[7] Xiao L, Honig B (1999) Electrostatic contributions to the stability of hyperthermophilic proteins. J Mol Biol 289: 1435–1444.

[8] Haney PJ, Badger JH, Buldak GL, Reich CI, Woese CR, et al. (1999) Thermal adaptation analyzed by comparison of protein sequences from mesophilic and extremely thermophilic methanococcus species. Proc Natl Acad Sci 96: 3578–3583.

[9] Beeby M, O'Connor BD, Ryttersgaard C, Boutz DR, Perry LJ, et al. (2005) The genomics of disulfide bonding and protein stabilization in thermophiles. PLoS Biol 3: e309.

[10] Jollivet D, Mary J, Gagnire N, Tanguy A, Fontanillas E, et al. (2012) Proteome adaptation to high temperatures in the ectothermic hydrothermal vent pompeii worm. PLoS ONE 7: e31150.

[11] Berezovsky IN, Shakhnovich EI (2005) Physics and evolution of thermophilic adaptation. Proc Natl Acad Sci 102: 12742–12747.

[12] Takano K, Aoi A, Koga Y, Kanaya S (2013) Evolvability of thermophilic proteins from archaea and bacteria. Biochemistry 52: 4774–4780.

[13] Gao J, Wang W (2012) Analysis of structural requirements for thermo-adaptation from orthologs in microbial genomes - springer. Annals of Microbiology 62: 1635–41.

[14] Siddiqui KS, Cavicchioli R (2006) Cold-adapted enzymes. Annu Rev Biochem 75: 403–433.

[15] Ayala-del Ro HL, Chain PS, Grzymski JJ, Ponder MA, Ivanova N, et al. (2010) The genome sequence of psychrobacter arcticus 273-4, a psychroactive siberian permafrost bacterium, reveals mechanisms for adaptation to low-temperature growth. Appl Environ Microbiol 76: 2304–2312.

[16] Casanueva A, Tuffin M, Cary C, Cowan DA (2010) Molecular adaptations to psychrophily: the impact of omic technologies. Trends Microbiol 18: 374–381.

[17] Daniel RM, Cowan DA, Morgan HW, Curran MP (1982) A correlation between protein thermostability and resistance to proteolysis. Biochem J 207: 641–644.

[18] Fontana A, Fassina G, Vita C, Dalzoppo D, Zamai M, et al. (1986) Correlation between sites of limited proteolysis and segmental mobility in thermolysin. Biochemistry 25: 1847–1851.

[19] Wrba A, Schweiger A, Schultes V, Jaenicke R, Zavodszky P (1990) Extremely thermostable d-glyceraldehyde-3-phosphate dehydrogenase from the eubacterium thermotoga maritima. Biochemistry 29: 7584–7592.

[20] Huyghues-Despointes BMP, Scholtz JM, Pace CN (1999) Protein conformational stabilities can be determined from hydrogen exchange rates. Nat Struct Mol Biol 6: 910–912.

[21] Zvodszky P, Kardos J, Svingor , Petsko GA (1998) Adjustment of conformational flexibility is a key event in the thermal adaptation of proteins. Proc Natl Acad Sci 95: 7406–7411.

[22] Eijsink VG, Bjrk A, Gseidnes S, Sirevg R, Synstad B, et al. (2004) Rational engineering of enzyme stability. J Biotechnol 113: 105–120.

[23] Reetz MT, Carballeira JD, Vogel A (2006) Iterative saturation mutagenesis on the basis of b factors as a strategy for increasing protein thermostability. Angew Chem Int Ed 45: 77457751.

[24] Becktel WJ, Schellman JA (1987) Protein stability curves. Biopolymers 26: 18591877.

[25] Cooper A (2010) Protein heat capacity: An anomaly that maybe never was. J Phys Chem Lett 1: 3298–3304.

[26] Nojima H, Ikai A, Oshima T, Noda H (1977) Reversible thermal unfolding of thermostable phosphoglycerate kinase. thermostability associated with mean zero enthalpy change. J Mol Biol 116: 429–442.

[27] Razvi A, Scholtz JM (2006) Lessons in stability from thermophilic proteins. Protein Sci 15: 15691578.

[28] Kumar S, Nussinov R (2004) Experiment-guided thermodynamic simulations on reversible two-state proteins: implications for protein thermostability. Biophys Chem 111: 235–246.

[29] Li Wt, Grayling RA, Sandman K, Edmondson S, Shriver JW, et al. (1998) Thermodynamic stability of archaeal histones. Biochemistry 37: 10563–10572.

[30] Hollien J, Marqusee S (1999) A thermodynamic comparison of mesophilic and thermophilic ribonucleases h. Biochemistry 38: 3831–3836.

[31] Gerday C, Aittaleb M, Bentahir M, Chessa JP, Claverie P, et al. (2000) Cold-adapted enzymes: from fundamentals to biotechnology. Trends Biotechnol 18: 103–107.

[32] Vieille C, Zeikus GJ (2001) Hyperthermophilic enzymes: Sources, uses, and molecular mechanisms for thermostability. Microbiol Mol Biol Rev 65: 1–43.

[33] Littlechild J, Novak H, James P, Sayer C (2013) Mechanisms of thermal stability adopted by thermophilic proteins and their use in white biotechnology. In: Satyanarayana T, Littlechild J, Kawarabayasi Y, editors, Thermophilic Microbes in Environmental and Industrial Biotechnology, Dordrecht: Springer Netherlands. pp. 481–507.

[34] Palmer A (2004) NMR characterization of the dynamics of biomacromolecules. Chem Rev 104: 3623–3640.

[35] Boehr DD, Dyson HJ, Wright PE (2006) An NMR perspective on enzyme dynamics. Chem Rev 106: 3055–3079.

[36] Kleckner IR, Foster MP (2011) An introduction to NMR-based approaches for measuring protein dynamics. Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics 1814: 942–968.

[37] Masterson LR, Cheng C, Yu T, Tonelli M, Kornev A, et al. (2010) Dynamics connect substrate recognition to catalysis in protein kinase a. Nat Chem Biol 6: 821–828.

[38] Tobi D, Bahar I (2005) Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state. Proc Natl Acad Sci 102: 18908–18913.

[39] Friedland GD, Lakomek NA, Griesinger C, Meiler J, Kortemme T (2009) A correspondence between solution-state dynamics of an individual protein and the sequence and conformational diversity of its family. PLoS Comput Biol 5: e1000393.

[40] Kurkcuoglu Z, Bakan A, Kocaman D, Bahar I, Doruker P (2012) Coupling between catalytic loop motions and enzyme global dynamics. PLoS Comput Biol 8: e1002705.

[41] Massi F, Wang C, Palmer AG (2006) Solution NMR and computer simulation studies of active site loop motion in triosephosphate isomerase. Biochemistry 45: 10787–10794.

[42] Shaw DE, Maragakis P, Lindorff-Larsen K, Piana S, Dror RO, et al. (2010) Atomic-level characterization of the structural dynamics of proteins. Science 330: 341–346.

[43] Long D, Brüschweiler R (2011) In silico elucidation of the recognition dynamics of ubiquitin. PLoS Comput Biol 7: e1002035.

[44] Xue Y, Ward JM, Yuwen T, Podkorytov IS, Skrynnikov NR (2012) Microsecond timescale conformational exchange in proteins: Using long molecular dynamics trajectory to simulate NMR relaxation dispersion data. J Am Chem Soc 134: 2555–2562.

[45] Pisliakov AV, Cao J, Kamerlin SCL, Warshel A (2009) Enzyme millisecond conformational dynamics do not catalyze the chemical step. Proc Natl Acad Sci 106: 17359–17364.

[46] Kamerlin SCL, Warshel A (2010) At the dawn of the 21st century: Is dynamics the missing link for understanding enzyme catalysis? Proteins : NA–NA.

[47] Bhabha G, Lee J, Ekiert DC, Gam J, Wilson IA, et al. (2011) A dynamic knockout reveals that conformational fluctuations influence the chemical step of enzyme catalysis. Science 332: 234–238.

[48] Liang ZX, Lee T, Resing KA, Ahn NG, Klinman JP (2004) Thermal-activated protein mobility and its correlation with catalysis in thermophilic alcohol dehydrogenase. Proc Natl Acad Sci 101: 9556–9561.

[49] Boekelheide N, Salomn-Ferrer R, Miller TF (2011) Dynamics and dissipation in enzyme catalysis. Proc Natl Acad Sci 108: 16159–16163.

[50] Adamczyk AJ, Cao J, Kamerlin SCL, Warshel A (2011) Catalysis by dihydrofolate reductase and other enzymes arises from electrostatic preorganization, not conformational motions. Proc Natl Acad Sci 108: 14115–14120.

[51] Loveridge EJ, Behiry EM, Guo J, Allemann RK (2012) Evidence that a "dynamic knockout" in Escherichia coli dihydrofolate reductase does not affect the chemical step of catalysis. Nature Chemistry 4: 292–297.

[52] Kohen A, Cannio R, Bartolucci S, Klinman JP, Klinman JP (1999) Enzyme dynamics and hydrogen tunnelling in a thermophilic alcohol dehydrogenase. Nature 399: 496–499.

[53] Wolf-Watz M, Thai V, Henzler-Wildman K, Hadjipavlou G, Eisenmesser EZ, et al. (2004) Linkage between dynamics and catalysis in a thermophilic-mesophilic enzyme pair. Nat Struct Mol Biol 11: 945–949.

[54] Henzler-Wildman KA, Lei M, Thai V, Kerns SJ, Karplus M, et al. (2007) A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. Nature 450: 913–916.

[55] Sikorski RS, Wang L, Markham KA, Rajagopalan PTR, Benkovic SJ, et al. (2004) Tunneling and coupled motion in the Escherichia coli dihydrofolate reductase catalysis. J Am Chem Soc 126: 4778–4779.

[56] Oyeyemi OA, Sours KM, Lee T, Kohen A, Resing KA, et al. (2011) Comparative Hydrogen-Deuterium exchange for a mesophilic vs thermophilic dihydrofolate reductase at 25 °C: identification of a single active site region with enhanced flexibility in the mesophilic protein. Biochemistry 50: 8251–8260.

[57] Bae E, Phillips GN (2004) Structures and analysis of highly homologous psychrophilic, mesophilic, and thermophilic adenylate kinases. J Biol Chem 279: 28202–28208.

[58] Couñago R, Wilson CJ, Peña MI, Wittung-Stafshede P, Shamoo Y (2008) An adaptive mutation in adenylate kinase that increases organismal fitness is linked to stability-activity trade-offs. Protein Eng Des Sel 21: 19–27.

[59] Somero GN (1978) Temperature adaptation of enzymes: Biological optimization through structure-function compromises. Annu Rev Ecol Syst 9: 1–29.

[60] Hernández G, LeMaster DM (2001) Reduced temperature dependence of collective conformational opening in a hyperthermophile rubredoxin. Biochemistry 40: 14384–14391.

[61] Fitter J, Heberle J (2000) Structural equilibrium fluctuations in mesophilic and thermophilic α-amylase. Biophys J 79: 1629–1636.

[62] Fitter J, Herrmann R, Dencher NA, Blume A, Hauss T (2001) Activity and stability of a thermostable α-amylase compared to its mesophilic homologue: Mechanisms of thermal adaptation. Biochemistry 40: 10723–10731.

[63] Fitter J (2005) Structural and dynamical features contributing to thermostability in α-amylases. Cell Mol Life Sci 62: 1925–1937.

[64] Cavagnero S, Debe DA, Zhou ZH, Adams MWW, Chan SI (1998) Kinetic role of electrostatic interactions in the unfolding of hyperthermophilic and mesophilic rubredoxins. Biochemistry 37: 3369–3376.

[65] Meinhold L, Clement D, Tehei M, Daniel R, Finney JL, et al. (2008) Protein dynamics and stability: The distribution of atomic fluctuations in thermophilic and mesophilic dihydrofolate reductase derived using elastic incoherent neutron scattering. Biophys J 94: 4812–4818.

[66] Roca M, Liu H, Messer B, Warshel A (2007) On the relationship between thermal stability and catalytic power of enzymes. Biochemistry 46: 15076–15088.

[67] Tadokoro T, Kanaya S (2009) Ribonuclease H: molecular diversities, substrate binding domains, and catalytic mechanism of the prokaryotic enzymes. FEBS Journal 276: 14821493.

[68] Kanaya S, Itaya M (1992) Expression, purification, and characterization of a recombinant ribonuclease H from Thermus thermophilus HB8. J Biol Chem 267: 10184–10192.

[69] Ishikawa K, Okumura M, Katayanagi K, Kimura S, Kanaya S, et al. (1993) Crystal structure of ribonuclease H from thermus thermophilus HB8 refined at 2.8åresolution. J Mol Biol 230: 529–542.

[70] Hollien J, Marqusee S (2002) Comparison of the folding processes of T. thermophilus and E. coli ribonucleases H. J Mol Biol 316: 327–340.

[71] Haruki M, Tanaka M, Motegi T, Tadokoro T, Koga Y, et al. (2007) Structural and thermodynamic analyses of Escherichia coli RNase HI variant with quintuple thermostabilizing mutations. FEBS Journal 274: 58155825.

[72] Butterwick JA, Patrick Loria J, Astrof NS, Kroenke CD, Cole R, et al. (2004) Multiple time scale backbone dynamics of homologous thermophilic and mesophilic ribonuclease HI enzymes. J Mol Biol 339: 855–871.

[73] Butterwick JA, Palmer AG (2006) An inserted Gly residue fine tunes dynamics between mesophilic and thermophilic ribonucleases H. Prot Sci 15: 2697–2707.

[74] Trbovic N (2008) A Unified View of Protein Dynamics: Integration of Molecular Dynamics Simulations and NMR Spectroscopy. Ph.D., Columbia University.

[75] Mahajan S, de Brevern AG, Offmann B, Srinivasan N (2013) Correlation between local structural dynamics of proteins inferred from NMR ensembles and evolutionary dynamics of homologues of known structure. J Biomol Struct Dyn 0: 1–8.

[76] Lai L, Yokota H, Hung LW, Kim R, Kim SH (2000) Crystal structure of archaeal RNase HII: a homologue of human major RNase H. Structure 8: 897–904.

[77] Cerritelli SM, Crouch RJ (2009) Ribonuclease H: the enzymes in eukaryotes. The FEBS journal 276: 1494–1505.

[78] Horiuchi T, Maki H, Sekiguchi M (1984) RNase H-defective mutants of Escherichia coli: A possible discriminatory role of RNase H in initiation of DNA replication. Molecular and General Genetics MGG 195: 17–22.

[79] Ogawa T, Okazaki T (1984) Function of RNase h in DNA replication revealed by RNase h defective mutants of escherichia coli. Molecular and General Genetics MGG 193: 231–237.

[80] Drolet M, Phoenix P, Menzel R, Mass E, Liu LF, et al. (1995) Overexpression of RNase H partially complements the growth defect of an Escherichia coli delta topA mutant: R-loop formation is a major problem in the absence of DNA topoisomerase I. Proc Natl Acad Sci 92: 3526–3530.

[81] Aguilera A, Garca-Muse T (2012) R loops: From transcription byproducts to threats to genome stability. Mol Cell 46: 115–124.

[82] Wahba L, Amon J, Koshland D, Vuica-Ross M (2011) RNase H and multiple RNA biogenesis factors cooperate to prevent RNA:DNA hybrids from generating genome instability. Mol Cell 44: 978–988.

[83] Helmrich A, Ballarino M, Tora L (2011) Collisions between replication and transcription complexes cause common fragile site instability at the longest human genes. Mol Cell 44: 966–977.

[84] Flicek P, Ahmed I, Amode MR, Barrell D, Beal K, et al. (2012) Ensembl 2013. Nucleic Acids Res 41: D48–D55.

[85] Frank P, Braunshofer-Reiter C, Wintersberger U (1998) Yeast RNase H(35) is the counterpart of the mammalian RNase HI, and is evolutionarily related to prokaryotic RNase HII. FEBS Letters 421: 23–26.

[86] Arudchandran A, Cerritelli S, Narimatsu S, Itaya M, Shin DY, et al. (2000) The absence of ribonuclease H1 or H2 alters the sensitivity of Saccharomyces cerevisiae to hydroxyurea, caffeine and ethyl methanesulphonate: implications for roles of RNases H in DNA replication and repair. Genes Cells 5: 789802.

[87] Filippov V, Filippova M, Gill S (2001) Drosophila RNase H1 is essential for development but not for proliferation. Mol Genet Genomics 265: 771–777.

[88] Cerritelli SM, Frolova EG, Feng C, Grinberg A, Love PE, et al. (2003) Failure to produce mitochondrial DNA results in embryonic lethality in Rnaseh1 null mice. Mol Cell 11: 807–815.

[89] Ruhanen H, Ushakov K, Yasukawa T (2011) Involvement of DNA ligase III and ribonuclease H1 in mitochondrial DNA replication in cultured human cells. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research 1813: 2000–2007.

[90] Gaidamakov SA, Gorshkova II, Schuck P, Steinbach PJ, Yamada H, et al. (2005) Eukaryotic RNases H1 act processively by interactions through the duplex RNA-binding domain. Nucleic Acids Res 33: 2166–2175.

[91] Nowotny M, Cerritelli SM, Ghirlando R, Gaidamakov SA, Crouch RJ, et al. (2008) Specific recognition of RNA/DNA hybrid and enhancement of human RNase H1 activity by HBD. The EMBO Journal 27: 1172–1181.

[92] Davies JF, Hostomska Z, Hostomsky Z, Jordan, Matthews DA (1991) Crystal structure of the ribonuclease H domain of HIV-1 reverse transcriptase. Science 252: 88–95.

[93] Katayanagi K, Miyagawa M, Matsushima M, Ishikawa M, Kanaya S, et al. (1990) Three-dimensional structure of ribonuclease H from E. coli. Nature 347: 306–309.

[94] Yang W, Hendrickson WA, Crouch RJ, Satow Y (1990) Structure of ribonuclease H phased at 2åresolution by MAD analysis of the selenomethionyl protein. Science 249: 1398–1405.

[95] Katayanagi K, Miyagawa M, Matsushima M, Ishikawa M, Kanaya S, et al. (1992) Structural details of ribonuclease H from Escherichia coli as refined to an atomic resolution. J Mol Biol 223: 1029–1052.

[96] Ishikawa K, Kimura S, Kanaya S, Morikawa K, Nakamura H (1993) Structural study of mutants of Escherichia coli ribonuclease HI with enhanced thermostability. Protein Eng 6: 85–91.

[97] Tadokoro T, Matsushita K, Abe Y, Rohman MS, Koga Y, et al. (2008) Remarkable stabilization of a psychrotrophic RNase HI by a combination of thermostabilizing mutations identified by the suppressor mutation method. Biochemistry 47: 8040–8047.

[98] Kimura S, Kanaya S, Nakamura H (1992) Thermostabilization of Escherichia coli ribonuclease HI by replacing left-handed helical lys95 with gly or asn. J Biol Chem 267: 22014–22017.

[99] Ratcliff K, Marqusee S (2010) Identification of residual structure in the unfolded state of ribonuclease h1 from the moderately thermophilic chlorobium tepidum: Comparison with thermophilic and mesophilic homologues. Biochemistry 49: 5167–5175.

[100] Tadokoro T, Kazama H, Koga Y, Takano K, Kanaya S (2013) Investigating the structural dependence of protein stabilization by amino acid substitution. Biochemistry 52: 2839–2847.

[101] Ratcliff K, Corn J, Marqusee S (2009) Structure, stability, and folding of ribonuclease H1 from the moderately thermophilic Chlorobium tepidum: Comparison with thermophilic and mesophilic homologues. Biochemistry 48: 5890–5898.

[102] Tadokoro T, You DJ, Abe Y, Chon H, Matsumura H, et al. (2007) Structural, thermodynamic, and mutational analyses of a psychrotrophic RNase HI,. Biochemistry 46: 7460–7468.

[103] Raschke TM, Kho J, Marqusee S (1999) Confirmation of the hierarchical folding of RNase H: a protein engineering study. Nat Struct Mol Biol 6: 825–831.

[104] Cecconi C, Shank EA, Bustamante C, Marqusee S (2005) Direct observation of the three-state folding of a single protein molecule. Science 309: 2057–2060.

[105] Dabora JM, Marqusee S (1994) Equilibrium unfolding of Escherichia coli ribonuclease H: Characterization of a partially folded state. Protein Sci 3: 14011408.

[106] Raschke TM, Marqusee S (1997) The kinetic folding intermediate of ribonuclease H resembles the acid molten globule and partially unfolded molecules detected under native conditions. Nat Struct Mol Biol 4: 298–304.

[107] Yamasaki K, Ogasahara K, Yutani K, Oobatake M, Kanaya S (1995) Folding pathway of escherichia coli ribonuclease HI: a circular dichroism, fluorescence, and NMR study. Biochemistry 34: 16552–16562.

[108] Chamberlain AK, Handel TM, Marqusee S (1996) Detection of rare partially folded molecules in equilibrium with the native conformation of RNaseH. Nat Struct Mol Biol 3: 782–787.

[109] Spudich GM, Miller EJ, Marqusee S (2004) Destabilization of the Escherichia coli RNase H kinetic intermediate: Switching between a two-state and three-state folding mechanism. J Mol Biol 335: 609–618.

[110] Hollien J, Marqusee S (1999) Structural distribution of stability in a thermophilic enzyme. Proc Natl Acad Sci 96: 13674–13678.

[111] Hu W, Walters BT, Kan ZY, Mayne L, Rosen LE, et al. (2013) Stepwise protein folding at near amino acid resolution by hydrogen exchange and mass spectrometry. Proc Natl Acad Sci 110: 7684–7689.

[112] Goedken ER, Marqusee S (2001) Native-state energetics of a thermostabilized variant of ribonuclease HI. J Mol Biol 314: 863–871.

[113] Robic S, Berger JM, Marqusee S (2002) Contributions of folding cores to the thermostabilities of two ribonucleases H. Prot Sci 11: 381–389.

[114] Robic S, Guzman-Casado M, Sanchez-Ruiz JM, Marqusee S (2003) Role of residual structure in the unfolded state of a thermophilic protein. Proc Natl Acad Sci 100: 11345–11349.

[115] Kanaya S, Oobatake M, Liu Y (1996) Thermal stability of Escherichia coli ribonuclease HI and its active site mutants in the presence and absence of the Mg2+ ion: Proposal for a novel catalytic role for Glu48. J Biol Chem 271: 32729–32736.

[116] Goedken ER, Keck JL, Berger JM, Marqusee S (2000) Divalent metal cofactor binding in the kinetic folding trajectory of Escherichia coli ribonuclease HI. Protein Sci 9: 19141921.

[117] Goedken ER, Marqusee S (1998) Folding the ribonuclease H domain of Moloney murine leukemia virus reverse transcriptase requires metal binding or a short N-terminal extension. Proteins 33: 135143.

[118] Nowotny M, Gaidamakov SA, Crouch RJ, Yang W (2005) Crystal structures of RNase H bound to an RNA/DNA hybrid: Substrate specificity and metal-dependent catalysis. Cell 121: 1005–1016.

[119] Nowotny M, Gaidamakov SA, Ghirlando R, Cerritelli SM, Crouch RJ, et al. (2007) Structure of human RNase H1 complexed with an RNA/DNA hybrid: Insight into HIV reverse transcription. Mol Cell 28: 264–276.

[120] Katayanagi K, Okumura M, Morikawa K (1993) Crystal structure of Escherichia coli RNase HI in complex with Mg2+ at 2.8å resolution: Proof for a single Mg2+-binding site. Proteins 17: 337346.

[121] Keck JL, Goedken ER, Marqusee S (1998) Activation/attenuation model for RNase H: A one-metal mechanism with second-metal inhibition. J Biol Chem 273: 34128–34133.

[122] Oda Y, Yoshida M, Kanaya S (1993) Role of histidine 124 in the catalytic function of ribonuclease HI from Escherichia coli. J Biol Chem 268: 88 –92.

[123] Goedken ER, Marqusee S (2001) Co-crystal of Escherichia coli RNase HI with Mn2+ ions reveals two divalent metals bound in the active site. J Biol Chem 276: 7266–7271.

[124] Tsunaka Y, Takano K, Matsumura H, Yamagata Y, Kanaya S (2005) Identification of single Mn2+ binding sites required for activation of the mutant proteins of E. coli RNase HI at glu48 and/or asp134 by X-ray crystallography. J Mol Biol 345: 1171–1183.

[125] Klumpp K, Hang JQ, Rajendran S, Yang Y, Derosier A, et al. (2003) Two-metal ion mechanism of RNA cleavage by HIV RNase H and mechanismbased design of selective HIV RNase H inhibitors. Nucleic Acids Res 31: 6852–6859.

[126] Nowotny M, Yang W (2006) Stepwise analyses of metal ions in RNase h catalysis from substrate destabilization to product release. The EMBO Journal 25: 1924–1933.

[127] Steitz TA, Steitz JA (1993) A general two-metal-ion mechanism for catalytic RNA. Proc Natl Acad Sci 90: 6498–6502.

[128] Rosta E, Nowotny M, Yang W, Hummer G (2011) Catalytic mechanism of RNA backbone cleavage by ribonuclease H from quantum Mechanics/Molecular mechanics simulations. J Am Chem Soc 133: 8934–8941.

[129] Elsässer B, Fels G (2010) Atomistic details of the associative phosphodiester cleavage in human ribonuclease h. Phys Chem Chem Phys 12: 11081.

[130] Zecchinon L, Claverie P, Collins T, D'Amico S, Delille D, et al. (2001) Did psychrophilic enzymes really win the challenge? Extremophiles 5: 313–321.

[131] Mandel AM, Akke M, Palmer I (1995) Backbone dynamics of Escherichia coli ribonuclease HI: Correlations with structure and function in an active enzyme. J Mol Biol 246: 144–163.

[132] Yamasaki K, Saito M, Oobatake M, Kanaya S (1995) Characterization of the internal motions of Escherichia coli ribonuclease HI by a combination of 15N-NMR relaxation analysis and molecular dynamics simulation: examination of dynamic models. Biochemistry 34: 6587–6601.

[133] Mandel AM, Akke M, Palmer AG (1996) Dynamics of ribonuclease H: Temperature dependence of motions on multiple time scales. Biochemistry 35: 16009–16023.

[134] Yamasaki K, Akasako-Furukawa A, Kanaya S (1998) Structural stability and internal motions of escherichia coli ribonuclease HI: 15N relaxation and hydrogen-deuterium exchange analyses. J Mol Biol 277: 707–722.

[135] Kroenke CD, Rance M, Palmer AG (1999) Variability of the 15N chemical shift anisotropy in escherichia coli ribonuclease h in solution. J Am Chem Soc 121: 10119–10125.

[136] Ishikawa K, Nakamura H, Morikawa K, Kimura S, Kanaya S (1993) Cooperative stabilization of Escherichia coli ribonuclease HI by insertion of Gly-80b and Gly-77 -¿ Ala substitution. Biochemistry 32: 7136–7142.

[137] Alder BJ, Wainwright TE (1959) Studies in molecular dynamics. i. general method. J Chem Phys 31: 459–466.

[138] Rahman A (1964) Correlations in the motion of atoms in liquid argon. Phys Rev 136: A405–A411.

[139] Swope WC, Andersen HC, Berens PH, Wilson KR (1982) A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. J Chem Phys 76: 637–649.

[140] Predescu C, Lippert RA, Eastwood MP, Ierardi D, Xu H, et al. (2012) Computationally efficient molecular dynamics integrators with improved sampling accuracy. Mol Phys 110: 967–983.

[141] Andersen HC (1980) Molecular dynamics simulations at constant pressure and/or temperature. J Chem Phys 72: 2384–2393.

[142] Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR (1984) Molecular dynamics with coupling to an external bath. J Chem Phys 81: 3684–3690.

[143] Nos S (1984) A unified formulation of the constant temperature molecular dynamics methods. J Chem Phys 81: 511–519.

[144] Hoover WG (1985) Canonical dynamics: Equilibrium phase-space distributions. Physical Review A 31: 1695–1697.

[145] Martyna GJ, Tobias DJ, Klein ML (1994) Constant pressure molecular dynamics algorithms. J Chem Phys 101: 4177–4189.

[146] Piana S, Lindorff-Larsen K, Dirks RM, Salmon JK, Dror RO, et al. (2012) Evaluating the effects of cutoffs and treatment of long-range electrostatics in protein folding simulations. PLoS ONE 7: e39918.

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

[148] Best RB, Hummer G (2009) Optimized molecular dynamics force fields applied to the HelixCoil transition of polypeptides. J Phys Chem B 113: 9004–9015.

[149] Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, et al. (2010) Improved side-chain torsion potentials for the amber ff99SB protein force field. Proteins : NA–NA.

[150] Li DW, Brüschweiler R (2011) Iterative optimization of molecular mechanics force fields from NMR data of full-length proteins. J Chem Theory Comput 7: 1773–1782.

[151] Best RB, Zhu X, Shim J, Lopes PEM, Mittal J, et al. (2012) Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone , and side-chain 1 and 2 dihedral angles. J Chem Theory Comput 8: 3257–3273.

[152] Mamatkulov S, Fyta M, Netz RR (2013) Force fields for divalent cations based on single-ion and ion-pair properties. J Chem Phys 138: 024505–024505-12.

[153] Halgren TA, Damm W (2001) Polarizable force fields. Curr Opin Struct Biol 11: 236–242.

[154] Ponder JW, Wu C, Ren P, Pande VS, Chodera JD, et al. (2010) Current status of the AMOEBA polarizable force field. J Phys Chem B 114: 2549–2564.

[155] Tuckerman M, Berne BJ, Martyna GJ (1992) Reversible multiple time scale molecular dynamics. J Chem Phys 97: 1990.

[156] Fan H, Mark AE (2003) Relative stability of protein structures determined by x-ray crystallography or NMR spectroscopy: A molecular dynamics simulation study. Proteins 53: 111120.

[157] Zeiske T, Stafford KA, Friesner RA, Palmer AG (2013) Starting-structure dependence of nanosecond timescale intersubstate transitions and reproducibility of MD-derived order parameters. Proteins 81: 499509.

[158] Mark P, Nilsson L (2001) Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. J Phys Chem A 105: 9954–9960.

[159] Wong V, Case DA (2008) Evaluating rotational diffusion from protein MD simulations. J Phys Chem B 112: 6013–6024.

[160] McCammon J, Gelin B, Karplus M (1977) Dynamics of folded proteins. Nature 267: 585–590.

[161] Karplus M, Mccammon JA (1979) Protein structural fluctuations during a period of 100 ps. Nature 277: 578–578.

[162] Lipari G, Szabo A, Levy RM (1982) Protein dynamics and NMR relaxation: comparison of simulations with experiment. Nature 300: 197–198.

[163] Allison JR (2012) Assessing and refining molecular dynamics simulations of proteins with nuclear magnetic resonance data. Biophysical Reviews 4: 189–203.

[164] Case DA (2002) Molecular dynamics and NMR spin relaxation in proteins. Acc Chem Res 35: 325–331.

[165] Karplus M, McCammon JA (2002) Molecular dynamics simulations of biomolecules. Nat Struct Mol Biol 9: 646–652.

[166] Lipari G, Szabo A (1980) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. theory and range of validity. J Am Chem Soc 104: 4546–4559.

[167] Maragakis P, Lindorff-Larsen K, Eastwood MP, Dror RO, Klepeis JL, et al. (2008) Microsecond molecular dynamics simulation shows effect of slow loop dynamics on backbone amide order parameters of proteins. J Phys Chem B 112: 6155–6158.

[168] Showalter SA, Brüschweiler R (2007) Validation of molecular dynamics simulations of biomolecules using NMR spin relaxation as benchmarks: application to the AMBER99SB force field. J Chem Theory Comput 3: 961–975.

[169] Long D, Li DW, Walter K, Griesinger C, Brüschweiler R (2011) Toward a predictive understanding of slow methyl group dynamics in proteins. Biophys J 101: 910–915.

[170] Yao L, Grishaev A, Cornilescu G, Bax A (2010) The impact of hydrogen bonding on amide 1H chemical shift anisotropy studied by cross-correlated relaxation and liquid crystal NMR spectroscopy. J Am Chem Soc 132: 10866–10875.

[171] Yao L, Grishaev A, Cornilescu G, Bax A (2010) Site-specific backbone amide 15N chemical shift anisotropy tensors in a small protein from liquid crystal and cross-correlated relaxation measurements. J Am Chem Soc 132: 4295–4309.

[172] Tang S, Case DA (2011) Calculation of chemical shift anisotropy in proteins. J Biomol NMR 51: 303–312.

[173] Wishart DS, Bigam CG, Holm A, Hodges RS, Sykes BD (1995) 1H, 13C and 15N random coil NMR chemical shifts of the common amino acids. i. investigations of nearest-neighbor effects. J Biomol NMR 5: 67–81.

[174] Wishart DS, Sykes BD, Richards FM (1992) The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. Biochemistry 31: 1647–1651.

[175] Shen Y, Bax A (2010) SPARTA+: a modest improvement in empirical NMR chemical shift prediction by means of an artificial neural network. J Biomol NMR 48: 13–22.

[176] Xu XP, Case DA (2001) Automated prediction of 15N, 13C, 13C and 13C chemical shifts in proteins using a density functional database. J Biomol NMR 21: 321–333.

[177] Neal S, Nip AM, Zhang H, Wishart DS (2003) Rapid and accurate calculation of protein 1H, 13C and 15N chemical shifts. J Biomol NMR 26: 215–240.

[178] Moon S, Case DA (2007) A new model for chemical shifts of amide hydrogens in proteins. J Biomol NMR 38: 139–150.

[179] Shen Y, Bax A (2007) Protein backbone chemical shifts predicted from searching a database for torsion angle and sequence homology. J Biomol NMR 38: 289–302.

[180] Kohlhoff KJ, Robustelli P, Cavalli A, Salvatella X, Vendruscolo M (2009) Fast and accurate predictions of protein NMR chemical shifts from interatomic distances. J Am Chem Soc 131: 13894–13895.

[181] Han B, Liu Y, Ginzinger SW, Wishart DS (2011) SHIFTX2: significantly improved protein chemical shift prediction. J Biomol NMR 50: 43–57.

[182] Markwick PRL, Cervantes CF, Abel BL, Komives EA, Blackledge M, et al. (2010) Enhanced conformational space sampling improves the prediction of chemical shifts in proteins. J Am Chem Soc 132: 1220–1221.

[183] Li DW, Brüschweiler R (2010) Certification of molecular dynamics trajectories with NMR chemical shifts. J Phys Chem Lett 1: 246–248.

[184] Robustelli P, Stafford KA, Palmer AG (2012) Interpreting protein structural dynamics from NMR chemical shifts. J Am Chem Soc 134: 6365–6374.

[185] Pietrucci F, Mollica L, Blackledge M (2013) Mapping the native conformational ensemble of proteins from a combination of simulations and experiments: New insight into the src-SH3 domain. J Phys Chem Lett 4: 1943–1948.

[186] Lindorff-Larsen K, Maragakis P, Piana S, Eastwood MP, Dror RO, et al. (2012) Systematic validation of protein force fields against experimental data. PLoS ONE 7: e32131.

[187] Beauchamp KA, Lin YS, Das R, Pande VS (2012) Are protein force fields getting better? a systematic benchmark on 524 diverse NMR measurements. J Chem Theory Comput 8: 1409–1414.

[188] Li DW, Brüschweiler R (2010) NMR-Based protein potentials. Angewandte Chemie 122: 69306932.

[189] Bae E, Phillips GN (2005) Identifying and engineering ion pairs in adenylate kinases: Insights from molecular dynamics simulations of thermophilic and mesophilic homologues. J Biol Chem 280: 30943–30948.

[190] Daily MD, Phillips GN, Cui Q (2011) Interconversion of functional motions between mesophilic and thermophilic adenylate kinases. PLoS Comput Biol 7: e1002103.

[191] Agarwal PK, Billeter SR, Rajagopalan PTR, Benkovic SJ, Hammes-Schiffer S (2002) Network of coupled promoting motions in enzyme catalysis. Proc Natl Acad Sci 99: 2794–2799.

[192] Pang J, Pu J, Gao J, Truhlar DG, Allemann RK (2006) Hydride transfer reaction catalyzed by hyperthermophilic dihydrofolate reductase is dominated by quantum mechanical tunneling and is promoted by both inter- and intramonomeric correlated motions. J Am Chem Soc 128: 8015–8023.

[193] Fan Y, Cembran A, Ma S, Gao J (2013) Connecting protein conformational dynamics with catalytic function as illustrated in dihydrofolate reductase. Biochemistry 52: 2036–2049.

[194] Bradley EA, Adams MW, Wampler JE, Stewart DE (1993) Investigations of the thermostability of rubredoxin models using molecular dynamics simulations. Protein Sci 2: 650665.

[195] Lazaridis T, Lee I, Karplus M (1997) Dynamics and unfolding pathways of a hyperthermophilic and a mesophilic rubredoxin. Protein Sci 6: 25892605.

[196] Grottesi A, Ceruso MA, Colosimo A, Di Nola A (2002) Molecular dynamics study of a hyperthermophilic and a mesophilic rubredoxin. Proteins 46: 287294.

[197] Okada J, Okamoto T, Mukaiyama A, Tadokoro T, You DJ, et al. (2010) Evolution and thermodynamics of the slow unfolding of hyperstable monomeric proteins. BMC Evolutionary Biology 10: 207.

[198] Rader AJ (2010) Thermostability in rubredoxin and its relationship to mechanical rigidity. Phys Biol 7: 016002.

[199] Philippopoulos M, Lim C (1995) Molecular dynamics simulation of E. coli ribonuclease H1 in solution: Correlation with NMR and X-ray data and insights into biological function. J Mol Biol 254: 771–792.

[200] Philippopoulos M, Lim C (1999) Exploring the dynamic information content of a protein NMR structure: Comparison of a molecular dynamics simulation with the NMR and X-ray structures of Escherichia coli ribonuclease HI. Proteins 36: 87–110.

[201] Fujiwara M, Kato T, Yamazaki T, Yamasaki K, Nagayama K (2000) NMR structure of ribonuclease HI from Escherichia coli. Biological & pharmaceutical bulletin 23: 1147–1152.

[202] Saito M, Tanimura R (1995) Relative melting temperatures of RNase HI mutant proteins from MD simulation/free energy calculations. Chem Phys Lett 236: 156–161.

[203] Tanimura R, Saito M (1996) Molecular Dynamics/Free energy perturbation studies of the thermostable V74I mutant of ribonuclease HI. Mol Simul 16: 75–85.

[204] Tang L, Liu H (2007) A comparative molecular dynamics study of thermophilic and mesophilic ribonuclease HI enzymes. J Biomol Struct Dyn 24: 379–392.

[205] Oda Y, Yamazaki T, Nagayama K, Kanaya S, Kuroda Y, et al. (1994) Individual ionization constants of all the carboxyl groups in ribonuclease HI from Escherichia coli determined by NMR. Biochemistry 33: 5275–5284.

[206] Anandakrishnan R, Aguilar B, Onufriev AV (2012) H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. Nucleic Acids Res .

[207] Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. J Chem Phys 79: 926–935.

[208] Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, et al. (2006) Comparison of multiple amber force fields and development of improved protein backbone parameters. Proteins 65: 712725.

[209] Himmel DM, Maegley KA, Pauly TA, Bauman JD, Das K, et al. (2009) Structure of HIV-1 reverse transcriptase with the inhibitor β-thujaplicinol bound at the RNase H active site. Structure 17: 1625–1635.

[210] Kirby KA, Marchand B, Ong YT, Ndongwe TP, Hachiya A, et al. (2012) Structural and inhibition studies of the RNase H function of xenotropic murine leukemia virus-related virus reverse transcriptase. Antimicrob Agents Chemother 56: 2048–2061.

[211] Zhou D, Chung S, Miller M, Le Grice SF, Wlodawer A (2012) Crystal structures of the reverse transcriptase-associated ribonuclease H domain of xenotropic murine leukemia-virus related virus. J Struct Biol 177: 638–645.

[212] Kim J, Kang S, Jung S, Yu K, Chung S, et al. (2012) Crystal structure of xenotropic murine leukaemia virus-related virus (XMRV) ribonuclease h. Biosci Rep 32: 455–463.

[213] Leo B, Schweimer K, Rsch P, Hartl MJ, Whrl BM (2012) The solution structure of the prototype foamy virus RNase H domain indicates an important role of the basic loop in substrate binding. Retrovirology 9: 73.

[214] Bowers KJ, Chow E, Xu H, Dror RO, Eastwood MP, et al. (2006) Scalable algorithms for molecular dynamics simulations on commodity clusters. In: Proceedings of the 2006 ACM/IEEE conference on Supercomputing. Tampa, Florida: ACM, p. 84. doi: 10.1145/1188455.1188544.

[215] Krutler V, van Gunsteren WF, Hnenberger PH (2001) A fast SHAKE algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations. J Comput Chem 22: 501508.

[216] Aaqvist J (1990) Ion-water interaction potentials derived from free energy perturbation simulations. J Phys Chem 94: 8021–8024.

[217] Chandrasekhar I, Clore G, Szabo A, Gronenborn AM, Brooks BR (1992) A 500 ps molecular dynamics simulation study of interleukin-1 in water: Correlation with nuclear magnetic resonance spectroscopy and crystallography. J Mol Biol 226: 239–250.

[218] Hunter S, Jones P, Mitchell A, Apweiler R, Attwood TK, et al. (2012) InterPro in 2011: new developments in the family and domain prediction database. Nucleic Acids Res 40: 4725–4725.

[219] Markowitz VM, Chen IMA, Palaniappan K, Chu K, Szeto E, et al. (2011) IMG: the integrated microbial genomes database and comparative analysis system. Nucleic Acids Res 40: D115–D122.

[220] Pei J, Kim BH, Grishin NV (2008) PROMALS3D: a tool for multiple protein sequence and structure alignments. Nucleic Acids Res 36: 2295–2300.

[221] Morcos F, Pagnani A, Lunt B, Bertolino A, Marks DS, et al. (2011) Direct-coupling analysis of residue coevolution captures native contacts across many protein families. Proc Natl Acad Sci 108: E1293–E1301.

[222] Duan Y, Wu C, Chowdhury S, Lee MC, Xiong G, et al. (2003) A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. J Comput Chem 24: 19992012.

[223] Cavalli A. almost. URL `www.open-almost.org`.

[224] Zhang H, Neal S, Wishart DS (2003) RefDB: a database of uniformly referenced protein chemical shifts. J Biomol NMR 25: 173–195.

[225] Delaglio F, Grzesiek S, Vuister G, Zhu G, Pfeifer J, et al. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR 6: 277–293.

[226] TD Goddard, Kneller D. Sparky. URL `http://www.cgl.ucsf.edu/home/sparky/`.

[227] Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, et al. (2005) The CCPN data model for NMR spectroscopy: Development of a software pipeline. Proteins 59: 687696.

[228] Helmus JJ, Jaroniec CP (2013) Nmrglue: an open source python package for the analysis of multidimensional NMR data. J Biomol NMR 55: 355–367.

[229] Keck JL, Marqusee S (1996) The putative substrate recognition loop of Escherichia coli ribonuclease H is not essential for activity. J Biol Chem 271: 19883–19887.

[230] Stahl SJ, Kaufman JD, Viki-Topi S, Crouch RJ, Wingfield PT (1994) Construction of an enzymatically active ribonuclease h domain of human immunodeficiency virus type 1 reverse transcriptase. Protein Eng 7: 1103–1108.

[231] Keck JL, Marqusee S (1995) Substitution of a highly basic helix/loop sequence into the RNase h domain of human immunodeficiency virus reverse transcriptase restores its mn(2+)-dependent RNase h activity. Proc Natl Acad Sci 92: 2740–2744.

[232] Kanaya S, Katsuda-Nakai C, Ikehara M (1991) Importance of the positive charge cluster in Escherichia coli ribonuclease HI for the effective binding of the substrate. J Biol Chem 266: 11621–11627.

[233] Henzler-Wildman KA, Thai V, Lei M, Ott M, Wolf-Watz M, et al. (2007) Intrinsic motions along an enzymatic reaction trajectory. Nature 450: 838–844.

[234] Pontiggia F, Zen A, Micheletti C (2008) Small- and large-scale conformational changes of adenylate kinase: A molecular dynamics study of the subdomain motion and mechanics. Biophys J 95: 5901–5912.

[235] Yamazaki T, Yoshida M, Kanaya S, Nakamura H, Nagayama K (1991) Assignments of backbone proton, carbon-13, and nitrogen-15 resonances and secondary structure of ribonuclease H from Escherichia coli by heteronuclear three-dimensional NMR spectroscopy. Biochemistry 30: 6036–6047.

[236] Trbovic N, Kim B, Friesner RA, Palmer AG (2008) Structural analysis of protein dynamics by MD simulations and NMR spin-relaxation. Proteins 71: 684–694.

[237] Hirano N, Haruki M, Morikawa M, Kanaya S (2000) Enhancement of the enzymatic activity of ribonuclease HI from thermus thermophilus HB8 with a suppressor mutation method. Biochemistry 39: 13285–13294.

[238] Nguyen TN, You DJ, Kanaya E, Koga Y, Kanaya S (2013) Crystal structure of metagenome-derived LC9-RNase h1 with atypical DEDN active site motif. FEBS Letters 587: 1418–1423.

[239] Kroenke C (2001) Dynamics of Escherichia coli ribonuclease H studied by 15N spin relaxation. Ph.D., Columbia University.

[240] Butterwick J (2006) On the dynamics of thermophilic proteins: Interrelationships between form, flexibility and function. Ph.D., Columbia University.

[241] Haruki M, Tsunaka Y, Morikawa M, Iwai S, Kanaya S (2000) Catalysis by Escherichia coli ribonuclease HI is facilitated by a phosphate group of the substrate. Biochemistry 39: 13939–13944.

[242] Bartlett GJ, Porter CT, Borkakoti N, Thornton JM (2002) Analysis of catalytic residues in enzyme active sites. J Mol Biol 324: 105–121.

[243] Trbovic N, Cho JH, Abel R, Friesner RA, Rance M, et al. (2009) Protein side-chain dynamics and residual conformational entropy. J Am Chem Soc 131: 615–622.

[244] Esadze A, Li DW, Wang T, Brüschweiler R, Iwahara J (2011) Dynamics of lysine side-chain amino groups in a protein studied by heteronuclear 1H15N NMR spectroscopy. J Am Chem Soc 133: 909–919.

[245] Best RB, Clarke J, Karplus M (2005) What contributions to protein side-chain dynamics are probed by NMR experiments? A molecular dynamics simulation analysis. J Mol Biol 349: 185–203.

[246] Showalter SA, Johnson E, Rance M, Brüschweiler R (2007) Toward quantitative interpretation of methyl side-chain dynamics from NMR by molecular dynamics simulations. J Am Chem Soc 129: 14146–14147.

[247] Ruschak AM, Kay LE (2010) Methyl groups as probes of supra-molecular structure, dynamics and function. J Biomol NMR 46: 75–87.

[248] Hansen AL, Lundstrm P, Velyvis A, Kay LE (2012) Quantifying millisecond exchange dynamics in proteins by CPMG relaxation dispersion NMR using side-chain 1H probes. J Am Chem Soc 134: 3178–3189.

[249] Pasat G, Zintsmaster JS, Peng JW (2008) Direct 13C-detection for carbonyl relaxation studies of protein dynamics. J Magn Reson 193: 226–232.

[250] Paquin R, Ferrage F, Mulder FAA, Akke M, Bodenhausen G (2008) Multiple-timescale dynamics of side-chain carboxyl and carbonyl groups in proteins by 13C nuclear spin relaxation. J Am Chem Soc 130: 15805–15807.

[251] Hansen AL, Kay LE (2011) Quantifying millisecond time-scale exchange in proteins by CPMG relaxation dispersion NMR spectroscopy of side-chain carbonyl groups. J Biomol NMR 50: 347–355.

[252] Kanaya S, Kohara A, Miura Y, Sekiguchi A, Iwai S, et al. (1990) Identification of the amino acid residues involved in an active site of Escherichia coli ribonuclease H by site-directed mutagenesis. J Biol Chem 265: 4615 –4621.

[253] Huang HW, Cowan JA (1994) Metallobiochemistry of the magnesium ion. Eur J Biochem 219: 253260.

[254] Tsunaka Y, Haruki M, Morikawa M, Oobatake M, Kanaya S (2003) Dispensability of glutamic acid 48 and aspartic acid 134 for Mn2+-dependent activity of Escherichia coli ribonuclease HI. Biochemistry 42: 3366–3374.

[255] Hostomsky Z, Hostomska Z, Matthews DA (1993) Ribonucleases h. In: Linn SM, Roberts RJ, editors, Nucleases, Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press. 2nd edition, pp. 341–376. URL `http://cshmonographs.org/csh/index.php/monographs/article/viewArticle/4158`.

[256] Oda Y, Nakamura H, Kanaya S, Ikehara M (1991) Binding of metal ions toE. coli RNase HI observed by1H15N heteronuclear 2D NMR. J Biomol NMR 1: 247–255.

[257] De Vivo M, Dal Peraro M, Klein ML (2008) Phosphodiester cleavage in ribonuclease H occurs via an associative two-metal-aided catalytic mechanism. J Am Chem Soc 130: 10955–10962.

[258] Rosta E, Woodcock HL, Brooks BR, Hummer G (2009) Artificial reaction coordinate tunneling in free-energy calculations: The catalytic reaction of RNase h. J Comput Chem 30: 1634–1641.

[259] Ho MH, De Vivo M, Dal Peraro M, Klein ML (2010) Understanding the effect of magnesium ion concentration on the catalytic activity of ribonuclease H through computation: Does a third metal binding site modulate endonuclease catalysis? J Am Chem Soc 132: 13702–13712.

[260] Fraczkiewicz R, Braun W (1998) Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules. J Comput Chem 19: 319333.

[261] Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. J Mol Graph 14: 33–38.

[262] Mal K, Barvk I (2013) Complex between human RNase HI and the phosphonate-DNA/RNA duplex: Molecular dynamics study. J Mol Graph Model 44: 81–90.

[263] Oda Y, Iwa S, Ohtsuka E, Ishikawa M, Ikehara M, et al. (1993) Binding of nucleic acids to E. coli RNase HI observed by NMR and CD spectroscopy. Nucleic Acids Res 21: 4690–4695.

[264] Klepeis JL, Lindorff-Larsen K, Dror RO, Shaw DE (2009) Long-timescale molecular dynamics simulations of protein structure and function. Curr Opin Struct Biol 19: 120–127.

[265] Plimpton S (1995) Fast parallel algorithms for short-range molecular dynamics. J Comput Phys 117: 1–19.

[266] Sugita Y, Okamoto Y (1999) Replica-exchange molecular dynamics method for protein folding. Chem Phys Lett 314: 141–151.

[267] Hamelberg D, Mongan J, McCammon JA (2004) Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules. J Chem Phys 120: 11919–11929.

[268] Mezei M (1987) Adaptive umbrella sampling: Self-consistent determination of the non-boltzmann bias. J Comput Phys 68: 237–248.

[269] Roux B (1995) The calculation of the potential of mean force using computer simulations. Comput Phys Commun 91: 275–282.

[270] Laio A, Gervasio FL (2008) Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. Reports on Progress in Physics 71: 126601.

[271] Barducci A, Bussi G, Parrinello M (2008) Well-tempered metadynamics: A smoothly converging and tunable free-energy method. Phys Rev Lett 100: 020603.

[272] Isralewitz B, Gao M, Schulten K (2001) Steered molecular dynamics and mechanical functions of proteins. Curr Opin Struct Biol 11: 224–230.

[273] Pande VS, Baker I, Chapman J, Elmer SP, Khaliq S, et al. (2003) Atomistic protein folding simulations on the submillisecond time scale using worldwide distributed computing. Biopolymers 68: 91109.

[274] Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, et al. (1983) CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. J Comput Chem 4: 187217.

[275] Weiner SJ, Kollman PA, Case DA, Singh UC, Ghio C, et al. (1984) A new force field for molecular mechanical simulation of nucleic acids and proteins. J Am Chem Soc 106: 765–784.

[276] Tsui V, Case DA (2000) Molecular dynamics simulations of nucleic acids with a generalized born solvation model. J Am Chem Soc 122: 2489–2498.

[277] Roux B, Simonson T (1999) Implicit solvent models. Biophys Chem 78: 1–20.

[278] Zhou R (2003) Free energy landscape of protein folding in water: Explicit vs. implicit solvent. Proteins 53: 148161.

[279] Fine R, Dimmler G, Levinthal C (1991) FASTRUN: a special purpose, hardwired computer for molecular simulation. Proteins 11: 242253.

[280] Toyoda S, Miyagawa H, Kitamura K, Amisaki T, Hashimoto E, et al. (1999) Development of MD engine: High-speed accelerator with parallel processor design for molecular dynamics simulations. J Comput Chem 20: 185199.

[281] Narumi T, Ohno Y, Okimoto N, Koishi T, Suenaga A, et al. (2006) A 55 TFLOPS simulation of amyloid-forming peptides from yeast prion sup35 with the special-purpose computer system MDGRAPE-3. In: Proceedings of the 2006 ACM/IEEE conference on Supercomputing. New York, NY, USA: ACM, SC '06.

[282] Kadau K, Germann TC, Lomdahl PS (2006) Molecular dynamics comes of age: 320 billion atom simulation on BlueGene/L. International Journal of Modern Physics C 17: 1755–1761.

[283] Shaw D, Dror R, Salmon J, Grossman J, Mackenzie K, et al. (2009) Millisecond-scale molecular dynamics simulations on anton. In: Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis. pp. 1–11. doi: 10.1145/1654059.1654126.

[284] Anderson JA, Lorenz CD, Travesset A (2008) General purpose molecular dynamics simulations fully implemented on graphics processing units. J Comput Phys 227: 5342–5359.

[285] Eastman P, Friedrichs MS, Chodera JD, Radmer RJ, Bruns CM, et al. (2013) OpenMM 4: A reusable, extensible, hardware independent library for high performance molecular simulation. J Chem Theory Comput 9: 461–469.

[286] Pierce LC, Salomon-Ferrer R, Augusto F de Oliveira C, McCammon JA, Walker RC (2012) Routine access to millisecond time scale events with accelerated molecular dynamics. J Chem Theory Comput 8: 2997–3002.

[287] Piana S, Lindorff-Larsen K, Shaw DE (2013) Atomistic description of the folding of a dimeric protein. J Phys Chem B .

[288] Piana S, Lindorff-Larsen K, Shaw D (2011) How robust are protein folding simulations with respect to force field parameterization? Biophys J 100: L47–L49.

[289] Ting D, Wang G, Shapovalov M, Mitra R, Jordan MI, et al. (2010) Neighbor-dependent ramachandran probability distributions of amino acids developed from a hierarchical dirichlet process model. PLoS Comput Biol 6: e1000763.

[290] Lim D, Gregorio GG, Bingman C, Martinez-Hackert E, Hendrickson WA, et al. (2006) Crystal structure of the moloney murine leukemia virus RNase h domain. J Virol 80: 8379–8389.

[291] Shan Y, Arkhipov A, Kim ET, Pan AC, Shaw DE (2013) Transitions to catalytically inactive conformations in EGFR kinase. Proc Natl Acad Sci 110: 7270–7275.

[292] Beach H, Cole R, Gill ML, Loria JP (2005) Conservation of μsms enzyme motions in the apo- and substrate-mimicked state. J Am Chem Soc 127: 9167–9176.

[293] Hanson JA, Duderstadt K, Watkins LP, Bhattacharyya S, Brokaw J, et al. (2007) Illuminating the mechanistic roles of enzyme conformational dynamics. Proc Natl Acad Sci 104: 18055–18060.

[294] Yuan Z, Zhao J, Wang ZX (2003) Flexibility analysis of enzyme active sites by crystallographic temperature factors. Protein Eng 16: 109–114.

[295] Kool ET (2002) Active site tightness and substrate fit in DNA replication. Annu Rev Biochem 71: 191–219.

[296] Honaker MT, Acchione M, Zhang W, Mannervik B, Atkins WM (2013) Enzymatic detoxication, conformational selection, and the role of molten globule active sites. J Biol Chem 288: 18599–18611.

[297] Haruki M, Noguchi E, Nakai C, Liu YY, Oobatake M, et al. (1994) Investigating the role of conserved residue Asp134 in Escherichia coli ribonuclease HI by site-directed random mutagenesis. Eur J Biochem 220: 623631.

[298] Tinberg CE, Khare SD, Dou J, Doyle L, Nelson JW, et al. (2013) Computational design of ligand-binding proteins with high affinity and selectivity. Nature 501: 212–216.