Dynamics of the P22 Bacteriophage I-domain by 15N NMR Relaxation and Native-State Hydrogen Exchange

LaTasha CR Fraser
University of Connecticut, lfras1@unh.newhaven.edu

Recommended Citation
Fraser, LaTasha CR, "Dynamics of the P22 Bacteriophage I-domain by 15N NMR Relaxation and Native-State Hydrogen Exchange" (2013). Master's Theses. 404.
https://opencommons.uconn.edu/gs_theses/404

This work is brought to you for free and open access by the University of Connecticut Graduate School at OpenCommons@UConn. It has been accepted for inclusion in Master's Theses by an authorized administrator of OpenCommons@UConn. For more information, please contact opencommons@uconn.edu.
Dynamics of the P22 Bacteriophage I-domain by $^{15}$N NMR Relaxation and Native-State Hydrogen Exchange

LaTasha Fraser
B.S., University of New Haven, 2011

M.S. Advisor: Andrei Alexandrescu

A Thesis
Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Science
At the
University of Connecticut
2013
Dynamics of the P22 Bacteriophage I-domain by $^{15}$N NMR Relaxation and Native-State Hydrogen Exchange

Presented by
LaTasha Fraser, B.S.

Major Advisor
__________________________
Andrei Alexandrescu

Associate Advisor
__________________________
Carolyn Teschke

Associate Advisor
__________________________
Arlene Albert

University of Connecticut
2013
Abstract

The bacteriophage P22 I-domain structure contains a 6-stranded β-barrel folding motif, a 3-stranded β-sheet, and a small α-helical region. This study describes the dynamics of the I-domain determined by $^{15}$N NMR relaxation and native-state hydrogen exchange experiments. Relaxation data suggests that the I-domain is experiencing the majority of its motion on the ps-ns timescale. $S^2$ order parameters show two loop regions, which are poorly defined in the NMR structure, are the most flexible regions. The thermodynamic stability of the I-domain to unfolding, as a function of denaturant, has been studied with circular dichroism (CD) and native-state hydrogen exchange (nsHX) experiments. The CD experiments show $\Delta G_{\text{spec}}$ values for unfolding to be about 6.2 kcal/mol at pH 6.5. By contrast, nsHX experiments showed a variety of $\Delta G_{\text{HX}}$ values ranging from 7.28 to 11.0 kcal/mol. This spread in stability indicates that the protein is undergoing sub-global unfolding, as opposed to global or local unfolding reactions. The β-barrel is shown to be the most stable region of the protein, and includes some super-stable residues, which have $\Delta G_{\text{HX}}$ values that are higher than the average of 8.3 kcal/mol. $\Delta G_{\text{HX}}$ values on average are about 2 kcal/mol higher than the estimated $\Delta G_{\text{spec}}$ value obtained from CD. This discrepancy can be due to the fact that nsHX experiments are measured at an individual residue level, and CD experiments monitor global unfolding transitions of the protein. $\Delta G_{\text{HX}}$ showed increased stability for the structured β-strand regions; however, the α-helical region of the I-domain showed no protection to exchange, indicating that this secondary structure is only marginally stable. Taken together, the relaxation and nsHX results suggest that the structurally conserved β-barrel core is the most stable and resistant to denaturation.
Introduction

Bacteriophage P22, a tailed double-stranded DNA virus that infects *Salmonella Enterica Serovar Typhimurium*, is comprised of a protein capsid made of hexamerically and pentamerically arranged coat proteins with a structure analogous to a bacteriophage HK97-fold (1). Bacteriophages are bacterial virions containing a DNA or RNA genome enclosed by a protein capsid. The P22 capsid has 415 copies of coat protein (CP) monomers, which form the capsid (1). CP interacts with scaffolding protein molecules and a portal protein complex to assemble a procapsid. The procapsid is a metastable intermediate that undergoes a maturation process during which scaffolding protein is released, while DNA is simultaneously packaged into the procapsid via the portal complex. DNA packaging results in a conformational change of the procapsid, leading to the formation of the mature viral capsid (2).

The P22 bacteriophage CP monomer has never been crystallized to obtain a high-resolution structure using x-ray crystallography. The related bacteriophage (HK97), however, has been crystallized and extensively studied. Cryo-electron microscopy (cryo-EM) reconstructions have established preliminary subnanometer-resolution structures of the P22 capsid. Using 8 Å cryo-EM data, Parent *et al.* proposed that the “extra-density” domain has a primarily β-sheet fold (1). Chen *et al.* performed a 3.8 Å cryo-EM reconstruction, and proposed that P22 had a core structure analogous to the HK97-fold with the addition of an “extra-density domain” (3). Although, both structures agreed on the β-sheet secondary structure, there was a disagreement involving residues 257-277, which were alternatively modeled as the D-loop region (3), or a β-strand (1). Additionally, one structure contained an α-helix (residues 322-326) (3), while the other did not (1). In order to resolve these differences, a high-resolution NMR structure of the I-domain was determined. The NMR structure showed that the I-domain is
markedly different from both cryo-EM reconstructions with a global RMSD of 10-17 Å. The
sequence of Chen’s proposed “D-loop” is actually a part of a structured β-sheet region, as
proposed by Parent. However, there is still a D-loop present, but the loop was incorrectly
threaded into the I-domain. In support of Chen’s structure, the I-domain does have an α-helical
region that is observed towards the C-terminal end of the sequence. Initially, the I-domain was
thought to be a telokin-like fold, which is a flat β-sandwich (1). However, the NMR structure
shows that the 6 β-strands are assembled into a β-barrel fold, instead of a β-sandwich. In
addition to the core β-barrel, there is also a small 3-stranded β-sheet region.

Currently, there are several theories about the function of the I-domain. Temperature-
sensitive folding mutants suggest the I-domain is the folding nucleus of the CP, and adds to the
stability of the CP monomers (1,2). By contrast, it has been suggested the I-domain makes inter-
capsomer contacts with the negative charges on the adjacent CP monomers acting as a stabilizer
between CPs (3).

This study investigates the backbone dynamics of the I-domain using NMR relaxation
and probes the stability of the I-domain due to unfolding via native-state hydrogen exchange
experiments (nsHX) as a function of the denaturant urea. Relaxation data show the residues
involved in normal hydrogen-bonded secondary structure are rigid on the ps-ns timescale. The
poorly defined disordered loop regions show faster relaxation rates indicating increased
flexibility on the ps-ns timescale. The protein does not show much increase in R\textsubscript{2,ex} parameters,
although some crosspeaks are weaker in the \textsuperscript{1}H-\textsuperscript{15}N HSQC spectrum due to fast hydrogen
exchange. The nsHX measurements give information on the free energy change due to hydrogen
exchange, ΔG\textsubscript{HX}, which indicates whether the protein is undergoing global, subglobal, or local
unfolding. As determined by nsHX, the β-barrel region of the I-domain shows the greatest
stability, with some residues exhibiting super-stability that is larger than the calculated average \( \Delta G_{\text{HX}} \) value. The edges of the \( \beta \)-strands in the \( \beta \)-barrel have lower stabilities. Furthermore, \( \beta_1 \) has the smallest \( \Delta G_{\text{HX}} \) values within the \( \beta \)-barrel, likely as a result of only interacting with one \( \beta \)-strand instead of interacting and being sandwiched between two \( \beta \)-stands. The small 3-stranded \( \beta \)-sheet outside the conserved \( \beta \)-barrel fold, has a weaker stability, while the short \( \alpha \)-helix has exchange rates that were too fast to measure, indicating marginal stability.

**Results**

**\( ^{15}\text{N} \) Relaxation Experiments**

Data for \( ^{15}\text{N} \) T\(_1\) (longitudinal) and T\(_2\) (transverse) relaxation times are shown in Figure 1. T\(_1\) is sensitive to motions in the ps-ns timescale, and T\(_2\) is sensitive to motions in the \( \mu \)s-ms timescale. The T\(_1\) and T\(_2\) relaxation time for each residue were determined by quantifying the loss in crosspeak intensity, within the \(^1\text{H}-^{15}\text{N} \) HSQC, for each of the 8 designated relaxation times. The values obtained from measuring the crosspeak intensities were then plotted on a semi-log scale, which illustrated an exponential decay curve, proportional to the time needed for each residue to completely relax. Representative residues from the rigid regions have slower relaxation times (e.g. D271, R325), when compared to flexible regions (e.g. T224, K249).
Figure 1. $T_1$ (A) and $T_2$ (B) data for representative residues. Blue data points are from residues in the rigid region of the protein and red from flexible residues. The relaxation data were fit to exponential decays, which are shown on a semi-logarithmic scale (y-axis).

The backbone dynamics of the I-domain are described by the relaxation rates $R_1$, $R_2$, and the $^1$H-$^{15}$N NOE. $R_1$ and $R_2$ relaxation rates were determined by calculating the inverse values of the $T_1$ and $T_2$ relaxation times, which represent the rate of HN relaxation. Motion observed above the average baseline for $R_1$ and $R_2$ could be due to exchange contributions or line-broadening from sampling more than one conformation. The $R_1$ relaxation rates are generally similar within the structured region of the protein and show increased values at the terminal ends, and the disordered loop regions between residues 239-255 and 281-291 (Figure 2A). The $R_2$ relaxation rates for the I-domain highlight the two loop regions (residues 239-255 and 281-291), which have a decreased relaxation rate when compared to the average rate 15.805 s$^{-1}$ (Figure 2B). The $R_2$ relaxation rates also show increased values for three residues experiencing line-broadening contributions due to fast amide proton exchange (residues A240, D246, and L281). $^1$H-$^{15}$N NOE values of the backbone amide protons for the I-domain are shown in figure 2C. The lower NOE values correspond to increased flexibility on the ns timescale for the terminal ends and two loop
regions of the protein. Excluding the flexible regions, all of the backbone $^1$H-$^{15}$N NOEs are at or above 0.7. The observed NOE mean of 0.736 is comparable to the theoretical maximum of 0.80 (14). This signifies that the regions of the protein involved in regular hydrogen-bonded secondary structure, are rigid on the ns timescale.

Figure 2. $^{15}$N NMR relaxation data as a function of residue position in the I-domain sequence. (A) $R_1$ relaxation rates. (B) $R_2$ relaxation rates. (C) $^{15}$N NOE values. Error bars (red) are shown for all amino acid residues but in some cases are smaller than the symbols used to represent the data. Secondary structure of the I-domain is indicated by boxes (α-helix) and arrows (β-sheet).
The backbone dynamics of the I-domain were analyzed using the ModelFree formalism using the program Tensor2 (Figure 3) (15). The rotational diffusion model calculated a correlation time of 11.0 ns for the I-domain. Figure 3A shows the $S^2$ order parameters, which represent the amplitude of HN bond motions on the ps-ns timescale, with a value of 1 indicating the rigid limit and 0 indicating the flexible limit. The lower $S^2$ order parameters correspond to the disordered loop regions in the NMR structure, which is also true for the N- and C- termini. Figure 3B illustrates the $R_{2,ex}$ terms, which show the exchange contributions determined by $R_2$ relaxation rates. $R_{2,ex}$ values are increased mostly toward the C-terminal end. Figure 4 shows the $S^2$ order parameters mapped out on a stereodiagram of the NMR structure bundle.

**Figure 3.** Order parameters (A) and $R_{2,ex}$ (B) obtained from the Model-Free calculations of the I-domain. The global correlation time for the I-domain is 11 ns. The secondary structure of the I-domain is indicated by boxes (α-helix) and arrows (β-sheet). Error bars (red) are shown for all amino acid residues but in some cases are smaller than the symbols used to represent the data.
Figure 4. Stereodiagram of the NMR structure bundle. The $\alpha$-helix is colored in blue, the $\beta$-sheet regions are colored in green, and the red regions are the two loops (239-255 and 281-291) and chain termini that show low $S^2$ order parameters in Figure 3A corresponding to the least ordered regions in the NMR structures.

Native-State Hydrogen Exchange

Prior to starting the native-state hydrogen exchange (nsHX) experiments, circular dichroism (CD) experiments were done to determine the stability of the I-domain as a function of pH. The stability of the I-domain decreased by about 3 kcal/mol at basic pH values. CD experiments showed that the I-domain is most stable between pH 6.0 and 7.6 (Figure 5A). Next, NMR HX experiments were performed at pH 6.0 and 7.0 to verify that the I-domain HN exchange was occurring dependent on the stability of the protein. The HN exchange was analyzed using the Linderstrom-Lang model:

\[
\text{“closed”} \quad \xrightarrow{k_{\text{op}}} \quad \text{“open”} \quad \xrightarrow{k_{\text{int}}} \quad \text{“exchanged”} \quad [1]
\]

where “closed” represents the exchange resistant conformation, “open” represents the exchange accessible conformation, $k_{\text{op}}$ is the unfolding rate of the structure, $k_{\text{cl}}$ is the folding rate of the
structure, \( k_{\text{int}} \) is the intrinsic chemical exchange rate, and “exchanged” represents the exchanged amide proton. The Linderstrom-Lang model states, when a protein is in the closed or native state, it undergoes a reversible opening reaction (\( k_{\text{op}} \)), which leads to the I-domain being in the open solvent accessible state. As long as the open conformation does not return to the closed conformation via the closing reaction (\( k_{\text{cl}} \)), the open conformation then undergoes an irreversible intrinsic chemical exchange, which results in the HN exchanging with the solvent. The observed exchange rate of the amide proton is equal to:

\[
k_{\text{obs}} = \frac{k_{\text{op}} k_{\text{int}}}{k_{\text{cl}} + k_{\text{int}}} \quad [2]
\]

The \( k_{\text{obs}} \) values can exchange within two limits. The first limit is defined as the EX1 limit, where:

\[
k_{\text{obs}} = k_{\text{op}} \quad [3]
\]

Under EX1 conditions, exchange becomes significantly faster than the closing rate, causing the exchange rates to become independent of the stability of the protein and dependent on unfolding rates; typically occurring at pH 10 or 11. The second limit \( k_{\text{obs}} \) values can exchange under is defined as the EX2 limit, where:

\[
k_{\text{obs}} = (k_{\text{op}}/k_{\text{cl}}) k_{\text{int}} \quad [4]
\]

Under EX2 conditions, the closing rate of the protein is significantly faster than the intrinsic chemical exchange rate, indicating the amide protons are exchanging dependent on the stability of the protein. When amides protons exchange dependent of stability or within the EX2 limit, every increase of one pH unit is the equivalent to a 10-fold increase in the hydrogen exchange.
rate. As opposed to EX1 conditions, the EX2 conditions typically occur at lower temperatures and pH values.

The conditions for HX have been investigated in the I-domain by plotting the log of the exchange rates at pH 6.0 and pH 7.0, for 0 M urea and 1.9 M urea, the largest urea concentration used for the nsHX experiments (Figure 5B, 5C). Theoretically, in the EX2 regime a log plot of exchange rates at pH 6.0 vs pH 7.0 should have a slope of about 1, and in the EX1 regime the same log plot should have a theoretical slope of about 0 (16). Figures 5B and 5C show linear regressions for the log of exchange values at pH 6.0 and 7.0. Exchange can shift from EX2 to EX1 with increasing urea concentrations, so the pH dependence was measured at 0 M and 1.9 M urea. The high concentration of urea can affect the slope due to the decrease in the stability of the structure, which is observed for 1.9 M urea (Figure 5C). Jaravine (2000) states, the decrease observed in the slope value indicates that the protein is approaching EX1 conditions, however, a slope higher than 0.7 is still consistent with EX2 conditions. In addition, the observed differences in the slopes for the samples in both 0 M and 1.9 M urea are close to unity as would be expected for a 10-fold increase in rates going from pH 6.0 to pH 7.0, and have been observed for other proteins such as cold shock protein A (17). The y-intercept is also an important factor in determining EX1 or EX2 conditions. Ideally, under EX2 conditions, the y-intercept should be inversely proportional to the difference in pH units between the tested pH values. At a urea concentration of 0 M, the I-domain had a y-intercept of -1.1 with an error of +/- 0.2, which is inversely proportional to the separation of one pH unit between pH 6.0 and 7.0. By contrast, the 1.9 M urea concentration exhibited a y-intercept of -1.5 with an error of +/- 0.1. Although the y-intercept for the 1.9 M urea exchange is lower than its predicted value for a pH difference of 1, the lower value is likely due to the decreased stability being experienced by the protein (16, 18,
19). However, because the y-intercept of the I-domain did not decrease to zero, the assumption can still be made that the protein is undergoing exchange within EX2 conditions. As previously stated, studies have shown that slopes above 0.7 are considered to be consistent with the EX2 conditions, because the linearity of the slope still indicates that the protein is experiencing closing rates that are faster than the exchange rates (16-19), thus verifying that even in the presence of 1.9 M urea the I-domain is undergoing exchange within EX2 conditions.

**Figure 5** I-domain $\Delta G_{\text{spec}}$ values obtained from circular dichroism experiments (222nm) as a function of pH dependence (A). Scatter plots of exchange rates on a logarithmic scale at 0 M urea (B) and 1.9 M urea (C) proving exchange is occurring under EX2 conditions.
I-domain Stability

Out of the 123 residues that comprise the I-domain, about 54 residues were detected in $^1$H-$^{15}$N HSQC experiments, after being dissolved in D$_2$O for an average of 40 minutes (Figure 6). Peaks that remained in the spectrum an hour after being dissolved were involved in the hydrogen-bonded secondary structure of the NMR models. Residues not involved in secondary structure exchanged too fast to be detected in the $^1$H-$^{15}$N HSQC.

Figure 6 Hydrogen exchange in the I-domain. Representative $^1$H-$^{15}$N HSQC spectra of the I-domain in water (A), after 44 min (B), after 3 days (C), and after 15 days (D) of exchange in D$_2$O at pH 6.0, in 0.9 M urea, at 25 °C. Slowly exchanging residues involved in secondary structure are labeled in blue (B,C,D).
Two forms of unfolding mechanisms have been proposed to explain how amide protons exchange with solvent (11-13). The first explanation involves “non-cooperative local fluctuations” of the native unfolded structure, which are hypothesized to be independent of denaturant. The second explanation involves “cooperative, partial or complete unfolding reactions” (10,20). These cooperative unfolding reactions are increasingly promoted as the concentration of denaturant (e.g. urea) is increased. Figure 7A illustrates CD unfolding denaturant curves for the I-domain, which show that the difference in free energy change as determined by CD ($\Delta G_{\text{spec}}$) is 6.2 with an error of +/- 0.3 kcal/mol. Figure 7B illustrates how exchange rates increase as a function of increasing denaturant concentration (Figure 7B).

**Figure 7** (A) Unfolding transitions for the I-domain monitored by circular dichroism at pH 6.0. $\Delta G_{\text{spec}}$ measured by CD is 6.2 (+/-0.3) kcal/mol. (B) Exponential decays of $^1$H-$^{15}$N crosspeak intensities due to denaturant concentration. Data are shown for a representative residue, S333.

The free energy change due to HX as a function of urea concentration, $\Delta G_{\text{HX}}$, was obtained for 53 slowly exchanging amide protons. The $\Delta G_{\text{HX}}$ values show a linear correlation for most of the residues, which is typical of cooperative exchange reactions (Figure 8).
Figure 8 Native-state hydrogen exchange of the I-domain. Representative plots of $\Delta G_{\text{HX}}$ values as a function of urea concentration. Error bars (black) are shown for all representative residues but in some cases are smaller than the symbols used to represent the data.

The standard-state free energy change for opening reactions in the absence of denaturant, $\Delta G_0$, was obtained by linear extrapolation of $\Delta G_{\text{HX}}$ values versus urea to zero denaturant concentration. Twelve residues with $\Delta G_{\text{HX}}$ values exhibited no slope indicative of amide proton exchange due to local fluctuations, defined as reactions that do not perturb a significant amount of surface area and are thus independent of denaturant (orange, Figure 9A).

The slope obtained from $\Delta G_{\text{HX}}$ values versus urea concentration yields $m_{\text{HX}}$, which gives information on the amount of change in accessible surface area that leads to amide proton exchange. The $\Delta G_0$, $\Delta G_f$, and $m_{\text{HX}}$ data for the I-domain are depicted in figures 9A and 9B. As previously stated, residues involved in secondary structure are shown to have the highest $\Delta G_0$ values. The higher $\Delta G_0$ values show that the $\beta$-strand regions are the more stable regions, which have an average $\Delta G_0$ of 8.3 kcal/mol. All 9 $\beta$-strands in the I-domain are represented in $^1$H-$^{15}$N HSQC spectra recorded in D$_2$O; however, the $\alpha$-helix (residues 322-326) seems to be completely unprotected, with exchange rates too fast to be measured. The highest $\Delta G_0$ values are found in
the β-barrel, showing that the region is highly cooperatively stabilized. The black horizontal line
in Figure 9A represents the ΔG\text{spec} values obtained from equilibrium unfolding transitions
monitored by circular dichroism (CD). The determined ΔG\text{spec} value of 6.2 kcal/mol is about 2.1
kcal/mol smaller than the average ΔG\text{0} value of 8.3 kcal/mol. Although there is a notable
difference between the ΔG\text{spec} and ΔG\text{0} values, this has been observed in other proteins (11, 21-
23).

A difference between ΔG\text{spec} and ΔG\text{0} has been observed for other proteins; such as RNase A. RNase A has a ΔG\text{0} value of 4.2 kcal/mol, which is 1.6 kcal/mol higher than its ΔG\text{spec} value (11). Another example is cytochrome C, which has a ΔG\text{0} value that is about 2 kcal/mol larger
than its ΔG\text{spec} value (11). There have been various hypothesized explanations for the
discrepancies between ΔG\text{0} and ΔG\text{spec} values. Bai (1994) proposes ΔG\text{0} can have an observed
increase due to any structure in the denatured state that may protect amide protons from
exchanging with deuterium. In addition, the stability of a protein has been reported to increase by
0.3 kcal/mol when dissolved in deuterium, as opposed to H₂O (12, 22). A second possible
discrepancy involves a non-linear dependence on denaturant concentration for ΔG\text{spec}. As the
denaturant approaches lower concentrations, the theoretical ΔG\text{spec} values undergo an upward
curvature. This causes an underestimation of observed ΔG\text{spec} values when extrapolated from
higher denaturant concentrations near the transition zone (23). In the case of ΔG\text{0}, the error in the
extrapolation is decreased because the intensities are measured far away from the transition zone;
at lower concentrations of denaturants (10). A third explanation for the discrepancy can be the
isomerization state of the prolines. The I-domain contains four prolines, one of which is cis.
When the four prolines are corrected for by using the method suggested by Bai (11), the proline
isomerization is predicted to contribute 1.12 kcal/mol to ΔG\text{spec} at 25 °C, increasing ΔG\text{spec} to 7.32
kcal/mol, and with D$_2$O contributions $\Delta G_{\text{spec}}$ is calculated at being 7.6 kcal/mol (Figure 9A, green horizontal line).

The residues at the edges of the $\beta$-barrel have average $\Delta G_0$ values and do not have the increased stability that is observed throughout the rest of the residues in the $\beta$-strands of the $\beta$-barrel. The 3-stranded $\beta$-sheet ($\beta_7$, $\beta_8$, and $\beta_9$) shows increased exchange rates and weaker protection, with the exception of V313, which has a strong $\Delta G_0$ value. There also seems to be some stability in the second major disordered loop at residues L286 and N287, which is consistent with the decreased flexibility observed in the $^1$H-$^15$N NOE and S$^2$ order parameters. These two regions do seem to be $\alpha$-helical in nature and are possibly as a result of residual structure left when the I-domain is disassociated from the coat protein. For $m_{\text{HX}}$ values, there seems to be the same distribution of surface accessibility, which is proportional to the extrapolated $\Delta G_0$ values.
Figure 9 (A) Thermodynamic free energy change from nsHX due to unfolding $\Delta G_0$ (blue) and local fluctuations $\Delta G_f$ (orange) reactions. The black horizontal line represents $\Delta G_{\text{spec}}$ from the unfolding transitions determined by circular dichroism, and the green horizontal line represents $\Delta G_{\text{spec}}$ values corrected for proline isomerization and D$_2$O contributions yielding a $\Delta G_{\text{spec}}$ value of 7.6 kcal/mol. (B) $m_{\text{HX}}$ values obtained from nsHX. Error bars (red) are shown for all amino acid residues. The secondary structure of the I-domain is indicated by boxes (α-helix) and arrows (β-sheet).
Figure 10 depicts the linear correlation observed between the $\Delta G_0$ and $m_{\text{HX}}$ values. In addition to the I-domain, this linear correlation has also been observed in other proteins such as T4 lysozyme, which leads to the likely assumption that the I-domain is undergoing subglobal unfolding reactions (19, 24). Subglobal unfolding should lead to co-variation of $\Delta G_0$ and $m_{\text{HX}}$, where the higher values are indicative of residues undergoing global unfolding reactions, and lower values represent partially folded sub-global unfolding reactions. The possibility of the linear correlation suggesting local fluctuation is ruled out because there would be no slope or correlation between $\Delta G_0$ and $m_{\text{HX}}$, due to the residues exchanging independent of denaturant. In addition, complete global unfolding of the protein can be ruled out, because in the unfolded state it assumed that all of the residues would have the same exchange rate, thus having the data points converge on similar values (19, 25).

![Figure 10](image-url)  
**Figure 10** Average extrapolated free energy change of opening reactions $\Delta G_0$ and $m_{\text{HX}}$ for 41 residues with exchange rates slow enough to be measured by NMR. Plot shows linear correlation indicating that the I-domain is undergoing sub-global unfolding.
Discussion

The P22 bacteriophage I-domain structure was previously characterized using low-resolution cryo-EM reconstructions (1,3). However, the two structures determined by each group are not in agreement. One cryo-EM structure proposed the I-domain is primarily β-sheet, with a short α-helix, and a protruding D-loop (3). The second alternative structure proposes that the protein is also primarily β-sheet; however, the “D-loop” is in a structured region within the β-sheet core (1). Determining the true structure of the I-domain is important because the structure of a protein correlates directly to its function. As a result of the differing proposed structures, there are differing suggestions for the function of the I-domain. Some viral coat proteins stabilize their capsids through cross-linked interactions with the neighboring CPs. However, P22 bacteriophage does not stabilize its protein capsid via cross-linking. Of the two differing structures, one structure suggests the I-domain CP monomers are stabilized by inter-capsomer binding with the “D-loop”, adding to the thermodynamic stability of the capsid (3). Conversely, the second suggestion for the function states that the I-domain serves a critical role in phage assembly, based on results obtained from temperature sensitive-folding mutants and their effects on the folding of the protein (1).

3D NMR experiments were performed on the I-domain, which characterized the high-resolution structure in solution. The NMR structure calculations determined that the structure for the I-domain is significantly different (10-15 Å) from both previous structures. The “D-loop” (residues 257-277), which has a high density of basic side-chains, was proposed to be important in making inter-capsomer contacts (3). However, the high-resolution NMR structure shows the “D-loop” region as being apart of a β-strand rather than a loop. In addition, the “D-loop” region is rigid based on the S² order parameters and NOE NMR relaxation data. S² order parameters
show that the most flexible regions of the I-domain on the ps-ns timescale are the two loops, 239-255 and 281-291. In addition, these loops show the poorest precision in the NMR structure of the I-domain. The flexibility of these loops could be intrinsic to the I-domain or could result from the I-domain being detached from the larger coat protein, if the loops govern the interaction between the I-domain and the rest of the P22 coat protein. Furthermore, there is an α-helix near the C-terminal region, as proposed by Chen (4), which is shown to be a rigid region. Currently, these possibilities are being tested by fitting the revised NMR structure into the electron density map.

There are relatively few line-broadening contributions due to \( R_{2,\text{ex}} \) processes, which would be indicative of motion on the \( \mu s-\text{ms} \) timescale. Based on pH titration data (not shown), some \(^1\text{H}-^{15}\text{N}\) correlations that correspond to residues with lowered peak intensities in the \(^1\text{H}-^{15}\text{N}\) HSQC spectrum, are attenuated due to fast amide proton exchange. Once the I-domain dynamics were determined by relaxation experiments, nsHX was used to understand the global and local stability of the protein. Figure 11 shows a structural mapping of the nsHX data on the I-domain NMR structure; with slowly exchanging residues highlighted according to their relative stabilities. The largest \( \Delta G_0 \) values were observed in the β-barrel with stability ranging from the average \( \Delta G_{\text{HX}} \) of 8.3 kcal/mol to values exhibiting super-protection with \( \Delta G_0 \) values as high as 8.6 kcal/mol. The strong protection observed in the β-barrel is, for the most part, consistent throughout the entire barrel, with the exception of the ends of each individual β-strand, which have smaller \( \Delta G_0 \) values. Another characteristic to note about the β-barrel are the residues with weaker stability in \( \beta_1 \). It is likely that these residues are weakly protected because they are only interacting with one β-stand, instead of being sandwiched between two adjacent β-strands. The structural data show that \( \beta_1 \) has two hydrogen bonds with \( \beta_2 \). Also, \( \beta_1 \) shows a kink at S231 and
G232, which likely causes the weak interactions, due to the misalignments between the two strands. The small 3-stranded sheet has residues with faster exchange rates, indicating weaker stability for that region, with the exception of strong protection for V313. The strong protection observed for V313 could be as a result of its close proximity to the second strand of the 3-stranded sheet and the \( \alpha \)-helix. Although residues with slow exchange rates are denaturant dependent, 12 residues had exhibited \( \Delta G_{\text{HX}} \) values with no linear correlation, indicating they are exchanging through local fluctuations. The observed \( \Delta G_f \) values due to local fluctuations can be a result of residues with low stability. \( \Delta G_f \) values typically have lower stabilities when compared to \( \Delta G_0 \) values. It is possible that once denaturation occurs, the already less stable residues undergo exchange with such small rates that they cannot be detected by nsHX experiments, thus showing \( \Delta G_{\text{HX}} \) values with no slope. Lastly, the grey segments depict regions of the protein where information could not be collected, because exchange occurred to quickly to be measured.

![Figure 11](image)

**Figure 11** Structural mapping of \( \Delta G_0 \) values. Regions with average \( \Delta G_0 \) values are colored in purple and the most stable regions, which are above the average \( \Delta G_0 \), are colored in pink. The least stable regions are colored in orange for weak and yellow for very weak. Residues experiencing \( \Delta G_f \) contributions only are colored in cyan. Grey regions indicate the residues for which no information is available; mostly due to these residues exchanging too fast to be measured.

The dynamics information obtained from this study, has lead to a more compete understanding of I-domain dynamics and the level of stability exhibited by the protein. The
HK97-fold of the P22 bacteriophage coat protein is highly conserved for not only bacterial viruses, but also eukaryotic virus such as HSV1, chicken pox, shingles, and epstein-barr. Studying the stability of the I-domain can offer information on the stabilities of the extra density domains for the eukaryotic viruses. The stability of the I-domain can be related to some implications for the virulence of the mature virus. The dynamics data was useful in confirming if the disordered loop regions were due to flexibility rather than an error in the structure calculations of the protein. Discovering the true function of the disordered loop regions is the next topic to be addressed for this protein, by docking the I-domain NMR structure onto the full-length capsid protein using the cryo-EM data. Ultimately, we would like to observe the interactions between the I-domain and the capsid in order to elucidate if the I-domain is indeed functioning as a folding nucleus or as a stabilizing unit for adjacent monomers. There are many further investigations of the I-domain, in terms of the full capsid protein, which are needed to determine the function of the entire protein. Studying the true functions of the flexible loop regions can lead to an understanding of their importance in comparison to the rest of the capsid. Mutation studies of the disordered loop regions can elucidate whether basic residues are necessary to maintain the interaction between the I-domain and the full capsid.

**Methods**

**I-domain expression and purification**

The I-domain was expressed in *E. coli* BL21(DE3) cells containing plasmid *pET30b* which encodes amino acids S223-V345 of the full-length coat protein(4). Kanamycin resistant cells were plated, selected, and grown at 37° C to mid-log phase with 40 µg/ml of kanamycin antibiotic in M9 minimal media. The expressed I-domain was induced with 1 mM isopropyl β-
D-1-thiogalactopyranoside (IPTG) for 16 h at 30° C. Isotopic labels were incorporated into the protein by adding 1 g/L of $^{15}$NH$_4$Cl to the minimal M9 media (Cambridge Isotopes, Andover, MA). After induction the cells were harvested via sedimentation at 8000 rpm in a Sorvall SLC-6000 rotor for 20 min. The cells were then resuspended in 20 mM sodium phosphate (pH 7.6) containing lysozyme (200 µg/ml), and a 1:100 dilution of EDTA-free protease inhibitor cocktail (Sigma, St. Louis, MO) to avoid protein degradation due to exogenous protease activity. The cells were lysed with a French press at 20,000 psi and treated at room temperature with 5 mM MgSO$_4$, 0.5 mM CaCl$_2$, and DNase and RNase (100 µg/ml each) for 30 min to degrade nucleic acids. Cell membrane fragments were removed by sedimentation in a Sorvall F18-12 x 50 rotor at 18,500 rpm for 30 min. The supernatant underwent ultracentrifugation at 40,000 rpm for 90 min to rid the sample of any lower molecular weight cell debris. The remaining supernatant was run through a Talon metal-affinity column (Clontech, Mountain View, CA) for purification of the I-domain. Fractions containing the I-domain were collected and run through a sizing column for additional purification. The protein was then precipitated with ammonium sulfate (NH$_4$SO$_4$) at 4° C, and the fraction between 45% and 57.5% NH$_4$SO$_4$ saturation was centrifuged in an SA-600 rotor for 15 min at 10,000 rpm. The remaining pellet was dialyzed three times in 20 mM sodium phosphate buffer (pH 6.0) and concentrated using a Centricon 10 K nominal molecular weight cut-off filter from Millipore (Billerica, MA).

$^{15}$N Relaxation Experiments

$T_1$, $T_2$, and NOE experiments were collected using a 1.5 mM, 300 µl sample, dissolved in 90% H$_2$O/10% D$_2$O. The sample was buffered to pH 6.06 with 20 mM sodium phosphate buffer containing 0.02% w/v sodium azide (NaN$_3$). NMR experiments were performed using a 600
MHz Varion INOVA spectrometer equipped with a cryogenic probe at 37° C. Longitudinal relaxation times ($T_1$) were measured using delays of 50, 130, 210, 490, 570, 710, and 990 ms. Transverse relaxation times ($T_2$) were determined using relaxation delays of 10, 30, 50, 70, 90, 110, 150, and 210 ms. Two separate $^1$H-$^{15}$N HSQC spectra were collected to obtain NOE values, one with proton saturation for 3 sec, the second, a control experiment, had the proton saturation period replaced by an equivalent 3 sec preacquisition delay. The data were processed using FELIX-NMR (San Diego, CA). Relaxation rates ($R_1$ and $R_2$) were calculated by fitting signal intensities as a function of relaxation time to a single exponential model:

$$I = I_0 \times \exp(-\tau/R_{1,2})$$  \[5\]

where $I$ is the intensity for the relaxation period $\tau$, $I_0$ is the initial amplitude, and $R_{1,2}$ represents either $R_1$ or $R_2$ relaxation rates (5). Experimental errors for the relaxation data were obtained from the standard errors of the fits. NOE values were calculated using the equation:

$$\text{NOE} = \frac{I(s)}{I(c)}$$  \[6\]

where $I(s)$ represents the intensities obtained from the saturated $^1$H-$^{15}$N HSQC and $I(c)$ represents the intensities obtained from the unsaturated control $^1$H-$^{15}$N HSQC. Errors for the NOE experiment were calculated using:

$$\text{NOE (error)} = \text{NOE} \times \sqrt{(r_0/(I(c)^2)) + (r_1/(I(s)^2))}$$  \[7\]

where $\text{NOE}$ is the value obtained from equation 2, $r_0$ is the root mean squared deviation (RMSD) baseline noise from the unsaturated spectrum, and $r_1$ is the RMSD of the baseline noise from the saturated spectrum. $R_1$, $R_2$, and NOE values were analyzed with the ModelFree
approach, assuming isotropic rotational diffusion; using the program Tensor2 (6). This allowed the characterization of internal motions in terms of $\tau_c$ (timescale of motion), $S^2$ order parameters (amplitude of motion), and $R_{2,ex}$ (exchange contributions).

**Circular Dichroism**

I-domain solutions were diluted to a final concentration of 0.2 mg/ml and mixed using a Hamilton Microlab 50 titrator to final urea concentrations ranging from 0 M to 6 M (7). Samples were incubated for 3 hours at 25° C, in the presence of 20 mM sodium phosphate at pH 4.5, 6.0, 7.6, 8.4, 9.0, and 11.0. The various pH values were necessary to determine the pH dependence of the I-domain’s stability to unfolding. In turn, CD experiments were also necessary to establish that the stability of the protein is constant between pH 6.0 and 7.0 to allow experiments that showed hydrogen exchange was occurring under the EX2 conditions for the nsHX experiments. The CD signal was read at 222 nm and averaged for 20 scans per sample with a slit width of 4 nm in a 2.0 mm pathlength cell. All samples were monitored with an Applied Photophysics (Leatherhead, Surrey, UK) Pi-Star 180 circular dichroism spectropolarimeter. $\Delta G_{\text{spec}}$ values were calculated using a linear least-squares fit of the CD titration curves.

**Native-State Hydrogen Exchange (nsHX) Experiments**

$^1\text{H}-^{15}\text{N}$ HSQC spectra were collected at 25 °C on a 600 MHz Varian Inova spectrometer equipped with a cryogenic probe. The data were processed using FELIX-NMR program (San Diego, CA). nsHX experiments were collected on samples with a final concentration of 0.17 mM in 20mM sodium phosphate buffer and 99.96% D$_2$O at an average pH of 6.64. Data were collected at 8 urea concentrations (0, 0.13, 0.23, 0.6, 0.8, 0.9, 1.7, and 1.9 M) determined by
refractometry (7). Exchange rates were obtained from intensities observed in two-dimensional 
\(^1\text{H}-^{15}\text{N}\) HSQC experiments collected with 1024 \(^1\text{H}\) x 100 \(^{15}\text{N}\) complex points and spectral widths of 7209.2 x 1404.3 Hz. Between 25 and 36 exchange time points were collected for each urea concentration, with experiment times ranging from 2 days at higher urea concentrations to 2.5 months at lower urea concentrations. Exchange rates for each residue were fit to a 3-parameter single-exponential decay:

\[ I = I_0 \exp(-k_{\text{obs}}t) + C \]  \[8\]

where \(I_0\) is the initial peak intensity, \(k_{\text{obs}}\) is the observed HX rate, and \(C\) is the baseline noise of the spectrum. The fits were done with KaleidaGraph version 3.6 (Synergy Software) (8). HX experiments at pH 6.02 and 7.01 were performed to prove exchange was occurring within the EX2 limit under the absence of urea and at the largest urea concentration used for the nsHX experiments (1.9 M). The free energy change of the open and closed conformation of the I-domain were calculated as:

\[ \Delta G_{\text{HX}} = -RT \ln(K_{\text{ex}}) = -RT \ln([\text{open}]/[\text{closed}]) = -RT \ln(k_{\text{obs}}/k_{\text{int}}) \]  \[9\]

where \(k_{\text{obs}}\) is the observed HX rate constant and \(k_{\text{int}}\) is the intrinsic HX rate constant estimated with the program SPHERE (9). The errors for \(\Delta G\) were propagated as previously described (10). The free energy difference due to HX, \(\Delta G_{\text{HX}}\), can have contributions from non-cooperative or cooperative reactions:

\[ \Delta G_{\text{HX}} = -RT \ln(\exp^{\Delta G_{\text{f,HX}}/RT} + \exp[(m_{\text{HX}} \times [\text{urea}]) - \Delta G_{u,\text{HX}}(0)]/RT) \]  \[10\]
where $\Delta G_{f,HX}$ is the change in free energy due to non-cooperative, local fluctuations, that are independent of denaturant concentrations (11-13). $\Delta G_{u,HX}(0)$ is the change in free energy due to cooperative partial or complete unfolding reactions, which are dependent on denaturant concentrations, and are extrapolated back to zero concentration of denaturant (11,13). The slope, $m_{HX}$, represents the amount of protein surface that is exposed to denaturant after unfolding occurs.
References


