

Atomic Resolution Structures of the Core Domain of Avian Sarcoma Virus Integrase and Its D64N Mutant^{†,‡}

Jacek Lubkowski,[§] Zbigniew Dauter,^{||} Fan Yang,[§] Jerry Alexandratos,[§] George Merkel,[⊥] Anna Marie Skalka,[⊥] and Alexander Wlodawer^{*,§}

Macromolecular Structure Laboratory, National Cancer Institute-Frederick Cancer Research and Development Center (NCI-FCRDC), ABL-Basic Research Program, Frederick, Maryland 21702, NCI-FCRDC and National Synchrotron Light Source, Brookhaven National Laboratory, Building 725A-X9, Upton, New York 11973, and Institute for Cancer Research, Fox Chase Cancer Center, Philadelphia, Pennsylvania 19111

Received June 15, 1999; Revised Manuscript Received August 5, 1999

ABSTRACT: Six crystal structures of the core domain of integrase (IN) from avian sarcoma virus (ASV) and its active-site derivative containing an Asp64 → Asn substitution have been solved at atomic resolution ranging 1.02–1.42 Å. The high-quality data provide new structural information about the active site of the enzyme and clarify previous inconsistencies in the description of this fragment. The very high resolution of the data and excellent quality of the refined models explain the dynamic properties of IN and the multiple conformations of its disordered residues. They also allow an accurate description of the solvent structure and help to locate other molecules bound to the enzyme. A detailed analysis of the flexible active-site region, in particular the loop formed by residues 144–154, suggests conformational changes which may be associated with substrate binding and enzymatic activity. The pH-dependent conformational changes of the active-site loop correlates with the pH vs activity profile observed for ASV IN.

Retroviruses, such as human immunodeficiency virus type 1 (HIV-1)¹ or avian sarcoma virus (ASV), encode in their genes three essential enzymes: reverse transcriptase (RT), protease (PR), and integrase (IN) (1). In rare cases a retrovirus such as feline immunodeficiency virus and equine infectious anemia virus also encodes dUTP-ase (2–4). The first three enzymes are considered to be primary targets for designing drugs against AIDS, because each enzyme is absolutely required for virus replication. Although the search for therapeutically suitable inhibitors has been successful for RT and PR and a number of drugs against these enzymes

are already in use (5), no such drugs targeted against IN are yet available. This is due in part to the lack of a complete structural description of IN, which is crucial for understanding its enzymatic activity. Such knowledge is the foundation of rational drug design (6).

A molecule of retroviral IN contains approximately 300 amino acids, and it comprises three domains: the zinc-binding N-terminal domain, the catalytic core domain, and the DNA-binding C-terminal domain. IN catalyzes the incorporation of reverse-transcribed viral DNA into the host genome in two steps, processing and joining (1, 7, 8), both of which involve nucleophilic attack by a hydroxyl group on a DNA phosphate. In the processing step, a water molecule attacks near the end of the viral DNA, displacing two nucleotides from the 3' ends of each of the viral DNA strands. In the joining step, each exposed viral DNA ribose 3'-OH is activated to attack the host DNA at a relatively nonspecific location with a separation of five or six nucleotides, thereby inserting viral DNA into the host genome. In vitro, these reactions require only IN, metal cations, and DNA, although other proteins play a role in vivo.

The structures of the N-terminal domain (9, 10) and C-terminal domain (11, 12) of HIV IN have been determined by nuclear magnetic resonance (NMR) spectroscopy, whereas the structure of the catalytic core domain has been solved by crystallography (13). The originally published structure of the catalytic core domain of HIV-1 IN lacked a number of residues in the active-site area (one catalytic residue and the entire flexible active-site loop) due to their disorder, and the conformations of the other side chains in the active site (most notably Asp64 and Glu116) diverged significantly from those found in the related enzymes (14–16). These distor-

[†] Research sponsored in part by the National Cancer Institute, DHHS, under contract with ABL. Other support includes National Institutes of Health grants CA-47486 and CA-06927, a grant for Infectious Disease Research from Bristol-Myers Squibb Foundation, and an appropriation from the Commonwealth of Pennsylvania. The contents of this publication do not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

[‡] In accordance with journal and NIH policy, coordinates of structures discussed in this article have been deposited with the Protein Data Bank, <http://www.rcsb.org/pdb/>, immediately available under the following codes: ASVIN-CIT, 1CXU; ASVIN-HEP, 1CZB; ASVIN-AS, 1CXQ; and D64N-CIT, 1CZ9.

^{*} To whom correspondence and reprint requests should be addressed. Phone: (301) 846-5036. Fax: (301) 846-6128. E-mail: wlodawer@ncifcrf.gov.

[§] National Cancer Institute-Frederick Cancer Research and Development Center. Phone: (301) 846-5036. Fax: (301) 846-6128.

^{||} NCI-FCRDC and National Synchrotron Light Source. Phone: (516) 244-5609.

[⊥] Institute for Cancer Research. Phone: (215) 728-2490. Fax: (215) 728-2778.

¹ Abbreviations: HIV-1, human immunodeficiency virus type 1; ASV, avian sarcoma virus; IN, integrase; rmsd, root-mean-square deviation.

tions of the active site were the most likely cause of the inability of crystallized HIV-1 IN to bind divalent cations. Recently, two new structures of HIV-1 IN were reported under different crystallization conditions (17, 18). In these structures, the conformation of acidic residues in the D,D-(35)E motif of the catalytic center was almost identical to that present in the ASV IN core, with a divalent metal cation complexed to the two aspartic acids (17) and, in one case, with an ordered active-site loop, as well (18). Modeling attempts to assemble a whole molecule of HIV-1 IN from the individual domains have been described (9, 19). However, the proposed models have not been verified so far, and ultimately they might not be very helpful for predictions of the enzymatic mechanism or for rational drug design.

Crystal structures of the catalytic core domain of ASV IN have also been published with divalent metal cations bound in the active site (15, 20), as well as with an inhibitor of HIV-1 IN (21). The active-site loop of ASV IN was shown to accommodate different conformations depending on the crystallization conditions (14, 20), although the most recent reports discussed difficulties in finding stable conformations of this loop (21, 22). Structural changes of the active site of ASV IN resulting from the Asp64 → Asn substitution or pH changes have also been reported recently (22), suggesting that the conformation of the flexible active-site loop of this enzyme could be correlated with pH. Aside from fundamental structural questions, such as the quaternary structure of the full-length enzyme or the structure of a complex between IN and DNA, even the description of isolated domains is still clearly incomplete and evolving.

To assess structural similarities and differences in ASV IN crystals grown under different conditions, we completed a series of studies of this enzyme at atomic resolution. In these experiments, we varied the precipitant, pH, and buffers, and we compared the active enzyme core domain with the inactive D64N derivative. These high-resolution, well-refined structures have now provided a consistent and more detailed picture of the active site of the enzyme's catalytic domain.

EXPERIMENTAL PROCEDURES

Preparation of Crystals. Crystals of the core domain of ASV IN were grown according to protocols described previously (14), using either PEG4000 or ammonium sulfate as the precipitant. In the case of the D64N derivative, only crystals grown in the presence of PEG4000 were studied. Crystals were grown at different pH values adjusted either with Hepes buffer (pH 7.5) or with citrate buffer (pH 6.0). To answer specific questions that arose during these studies, we transferred some crystals from their original solutions to different mother liquors before collecting X-ray data. Several crystals grown in citrate buffer were transferred stepwise to an acetate buffer-based mother liquor at an equivalent pH, but additionally containing MnCl₂ (50 mM). For the D64N derivative, a few crystals grown in citrate buffer were transferred to the Hepes buffer-based mother liquor (pH 7.5), because crystals of this protein could not be grown in Hepes buffer. Six different types of crystals were investigated in this study (Table 1).

Data Collection and Structure Refinement. The diffraction data for all crystals were collected on a synchrotron source (beamline X9B at NSLS, Brookhaven National Laboratory).

All experiments were performed on crystals rapidly frozen in a stream of nitrogen gas (temperature 95 K, Oxford Cryosystems) after immersing them for a few seconds in the cryoprotectant solution consisting of 85% of the appropriate crystal growth mother liquor and 15% glycerol. The intensities were recorded on a MAR345 image plate detector at a wavelength of 0.98 Å or on an ADSC four element CCD detector at 1.05 Å (data sets ASVIN-TRN and D64N-TRN; for an explanation of the abbreviations, see Table 1). For each crystal, X-ray data were collected in several passes (using different exposure times for each pass) and were merged together, resulting in very complete data sets. All data were processed with the software package HKL2000 (23). Data collection statistics are shown in Table 1.

Most of these structures, with the exception of ASVIN-TRN and D64N-TRN, were described previously at lower resolution (14, 22). Therefore, only rigid body refinement was necessary for obtaining starting models. The first structure refined in this series was ASVIN-CIT, based on our previous structure (14) (PDB accession code 1asv). Before rigid body refinement, all solvent molecules and residues 144–152 (forming part of the flexible active site loop) were removed from the model, and the side chain of Asp 64 was truncated beyond Cβ. In our previous reports (21, 22), we noted a new conformation of the active-site loop, as well as vastly different conformations of the side chain of Asp64; therefore, in the present study, we sought to minimize model bias. The refinement was cross-validated by the R_{free} index (24), which was calculated using 10% of all reflections.

The refinement of the ASVIN-CIT structure was carried out with the program X-PLOR (25), using the energy function as well as X-ray terms as targets in the minimization procedure. Rigid body refinement in the resolution range 3.0–8.0 Å was performed to compensate for any small differences in the unit cell parameters. This refinement was followed by positional and overall *B*-factor refinements. After visual inspection of the refined structures using the program O (26) and manual corrections, the resolution was subsequently extended and *B* factors were individually refined for all non-hydrogen atoms. When the X-ray data were extended to the highest resolution available, the refinement was continued using the program SHELXL (27). The quality of the geometrical and stereochemical indices was continuously monitored using the program PROCHECK (28). The refinement was also performed with the program REFMAC (29), using the maximum-likelihood target function. Because the results were consistent, this method of refinement was discontinued.

A different approach was used for the refinement of three other structures: ASVIN-AS, ASVIN-HEP, and D64N-CIT. For each, the final structure of ASVIN-CIT described above was used as the starting model. All solvent molecules were removed, and the model was subjected to isotropic refinement using the programs REFMAC and ARP (30). The latter program was used to select water molecules on the basis of $2F_o - F_c$ and $F_o - F_c$ maps. The refinement of these three models was continued by using the program SHELXL against F^2 with the conjugate gradient (CGLS) option. During the first five cycles, isotropic *B* factors were refined for all atoms. Later, non-hydrogen atoms were refined using anisotropic displacement parameters. At this stage, hydrogen

Table 1: Crystallization Conditions, Data Collection, and Refinement Statistics for ASV IN Structures

selected crystallization conditions	ASVIN-CIT	ASVIN-HEP	ASVIN-AS	D64N-CIT	ASVIN-TRN	D64N-TRN
buffer (pH)	0.1 M citrate (6.0)	0.1 M Hepes (7.5)	0.1 M Hepes (7.5)	0.1 M citrate (6.0)	0.1 M acetate (6.0)	0.1 M Hepes (7.5)
precipitant	20% PEG4000	20% PEG4000	2 M (NH ₄) ₂ SO ₄	20% PEG4000	20% PEG4000	20% PEG4000
protein definition	core	core	core	core, mutant D64N	core	core, mutant D64N
Data Collection Statistics						
total no. of reflections	357 298	250 390	1 034 785	345 764	262 703	345 283
no. of unique reflections	34 367	69 419	89 063	53 915	58 386	59 166
resolution range (Å)	13.0–1.42	15.0–1.06	20.0–1.02	20.0–1.20	15.0–1.15	15.0–1.15
completeness (overall) (%)	99.0	90.1	100.0	97.1	93.5	94.7
completeness (highest resolution shell) (%)	98.9 (1.47)	86.7 (1.08)	99.6 (1.04)	90.4 (1.22)	70.7 (1.19)	81.9 (1.19)
R_{sym}	0.046	0.034	0.056	0.060	0.037	0.046
average $I/\sigma(I)$ (overall)	39.5	32.3	43.6	28.4	35.5	44.8
average $I/\sigma(I)$ (highest resolution shell)	5.2	3.2	2.9	2.2	2.0	9.5
Data Collection Strategy						
no. of passes	2	3	3	2	2	2
exposure time/frame (s)	5/150	12/20/120	15/30/120	60/180	20/60	20/60
maximum resolution (Å)	2.0/1.42	2.40/1.40/1.06	2.35/1.50/1.02	2.40/1.20	2.50/1.15	2.50/1.15
unit cell (a/c , Å)	66.18/78.73	65.37/80.14	65.54/80.12	66.56/78.36	66.13/78.59	65.54/80.07
Refinement Statistics						
resolution range (Å)	10.0–1.42		10.0–1.06		10.0–1.02	10.0–1.20
crystallographic R^a	0.187		0.143		0.129	0.147
rmsd bonds (Å)	0.014		0.016		0.015	0.020
angle distances (Å)	0.029		0.031		0.030	0.029
Avg Coordinate Errors (Å) (No. of Atoms) ^b						
all atoms			0.020 (1034)		0.019 (1031)	0.028 (1030)
C			0.021 (652)		0.020 (648)	0.029 (645)
O			0.018 (184)		0.016 (189)	0.026 (197)
N			0.021 (194)		0.020 (192)	0.028 (184)
S			0.008 (4)		0.006 (3)	0.012 (4)
Average B Factor (Å ²)						
all protein non-H atoms	19.9		15.2		13.2	18.4
main-chain atoms	17.5		13.5		11.5	16.2
side-chain atoms	22.4		16.9		14.9	20.7
solvent atoms	36.8		20.0		26.7	27.3
no. of non-hydrogen protein atoms	1118		1132		1155	1114
no. of solvent atoms	203		137		192	211
no. of heterogen atoms ^c	19		15		27	18

^a Crystallographic R values are calculated on the basis of all observed intensities, which were measured within the resolution range used during the refinement. ^b The average errors of coordinates are obtained by inversion of the LSQ matrix. Only the atoms in single sites are included (disordered sites excluded; number of included atoms shown in parentheses). ^c The number shown corresponds to the nonprotein, nonsolvent atoms (citrate and glycerol molecules).

atoms were introduced into the well-ordered parts of the structure at stereochemically calculated positions. The positions of hydrogen atoms were correlated with those of bonded heavier atoms. For all hydrogen atoms included in the refinement, isotropic B factors that were 20% higher than those of the parent atoms (50% higher in the case of methyl hydrogens) were applied. Manual adjustments of the models were performed with the program Quanta (Molecular Simulations, Inc.). The occupancies of atoms present in double conformations were refined as constrained to x and $(1 - x)$. During the course of refinement, standard restraints recommended by Sheldrick and Schneider (27) were applied to the positional and displacement parameters, whereas geometrical restraints were those of Eng and Huber (31). The bulk solvent contribution to the structure factors was calculated with the SWAT option of SHELXL. Water molecules were classified as either fully or half-occupied on the basis of their electron density and their distances to neighboring atoms. Their occupancies were not refined,

because refinement of both occupancies and temperature factors at resolutions approaching 1 Å is generally not stable (32). The progress of refinement was monitored using the R_{free} index based on a reserved set (2%) of reflections. All data were used in the last refinement cycle, and the models were refined with the blocked full-matrix least-squares option. The blocks consisted of positional and anisotropic thermal parameters of 12 successive residues and overlapped by two residues. This approach allowed the individual uncertainties associated with all refined parameters to be properly estimated from the inversion of the least-squares matrix.

The structures of ASVIN-TRN and D64N-TRN were only partially refined using the program X-PLOR. As the starting models, the final structures of ASVIN-CIT and D64N-CIT, respectively, were used, omitting the side chains of residue 64, residues 144–152, and the solvent. Although some minor structural changes were observed in both cases, the conformations were not novel, and thus, the refinement was not

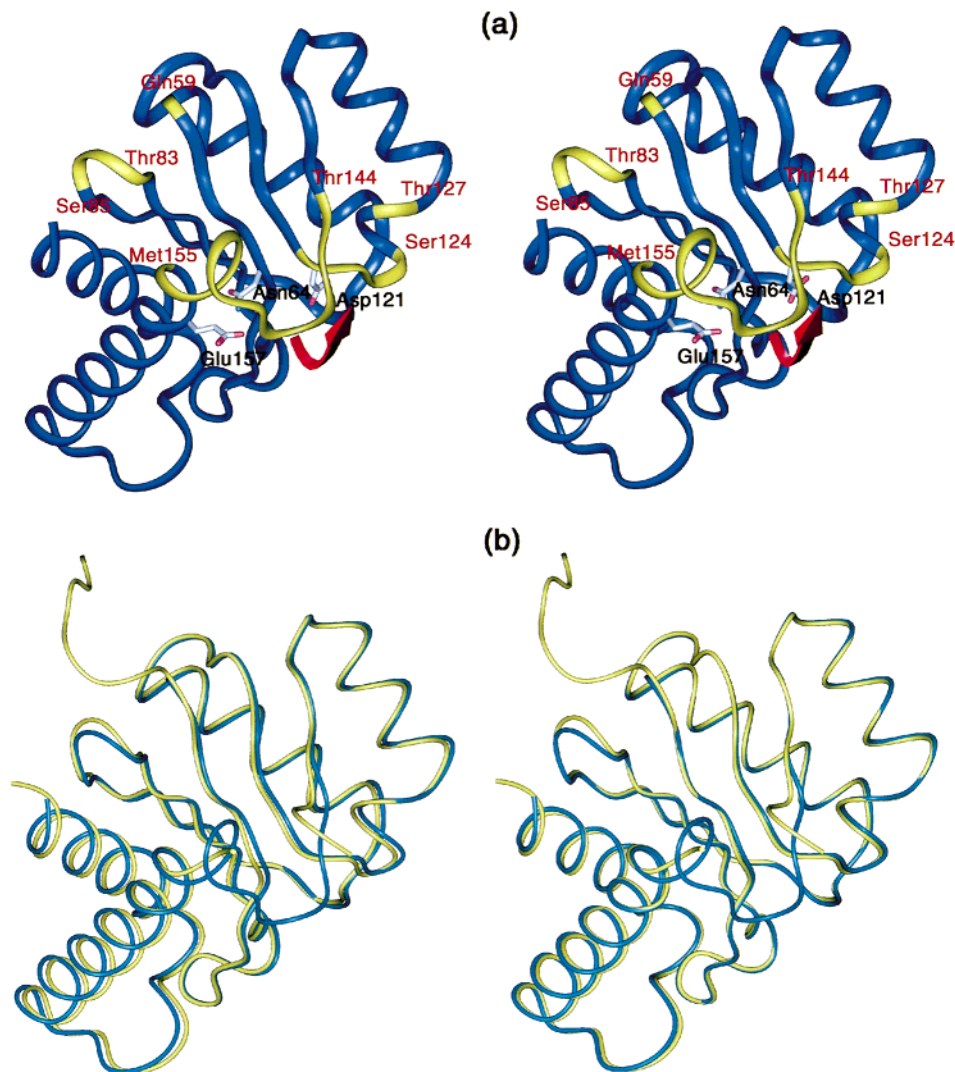


FIGURE 1: Stereo representations of the core domain of ASV IN. (A) Ribbon diagram of D64N-CIT, colored according to the isotropic B-factors of C α atoms. The regions of the enzyme consisting of C α atoms with B factors lower than 18 Å² are shown with a blue ribbon while the more flexible regions are highlighted in yellow. There are five regions of the enzyme characterized by higher mobility, and these regions are labeled in red. The longest flexible region comprises the active-site loop, Thr144-Met155. The red arrow indicates the direction of loop motion as determined by analysis of the anisotropic displacement factors. Three side chains of the active site residues are also shown and labeled in black. (B) Ribbon diagrams of two superimposed structures, ASVIN-AS (yellow) and D64N-CIT (blue), oriented identically to the model shown in panel A. The most flexible fragments of the enzyme, highlighted in yellow in panel A, have the poorest structural agreement between the two structures. Again, the discrepancies are largest for the flexible active site loop; in ASVIN-AS, it is completely disordered, and thus undetermined.

continued. In Table 1, the statistics and quality assessments for the first four structures are shown, whereas only experimental data are listed for ASVIN-TRN and D64N-TRN.

RESULTS

The atomic resolution models of ASV IN presented here provide a number of new structural details and an overall unparalleled accuracy in terms of both atomic position and thermal motion description, compared with the previous reports (14, 15, 20, 22). Because the structure of ASV IN under different crystallization conditions has been described previously, only the new observations will be discussed here in detail.

Comparison of the Structures of ASV IN. The most striking differences among the various structures are the changes in the flexibility of the loop (residues 144–154), caused by variation in pH. Most of the amino acids in this loop do not have an ordered structure at the “high” pH value (i.e., pH

7.5; ASVIN-HEP and ASVIN-AS) but are ordered at the “low” value (pH 6.0). Under the low-pH conditions, we could not determine the orientation of four (ASVIN-CIT) or seven (D64N-CIT) residues at the N-terminus, although they were clearly visible in both of the high pH structures (Figure 1). To assess the overall similarity among the different models, we aligned the four IN molecules on the basis of their C α atoms, using the program ALIGN (33). These alignments resulted in root-mean-square deviation (rmsd) values of 0.19 Å for ASVIN-CIT vs D64N-CIT (132 C α atoms used), 0.10 Å for ASVIN-HEP vs ASVIN-AS (for 136 C α atoms), 0.25 Å for ASVIN-AS vs D64N-CIT (for 126 C α atoms), and 0.24 Å for ASVIN-HEP vs D64N-CIT (for 128 C α atoms). Figure 2B shows the distribution of discrepancies among the positions of equivalent C α atoms obtained after superposition of the different structures.

We found the two high pH models to be remarkably similar. The largest deviation between the positions of two

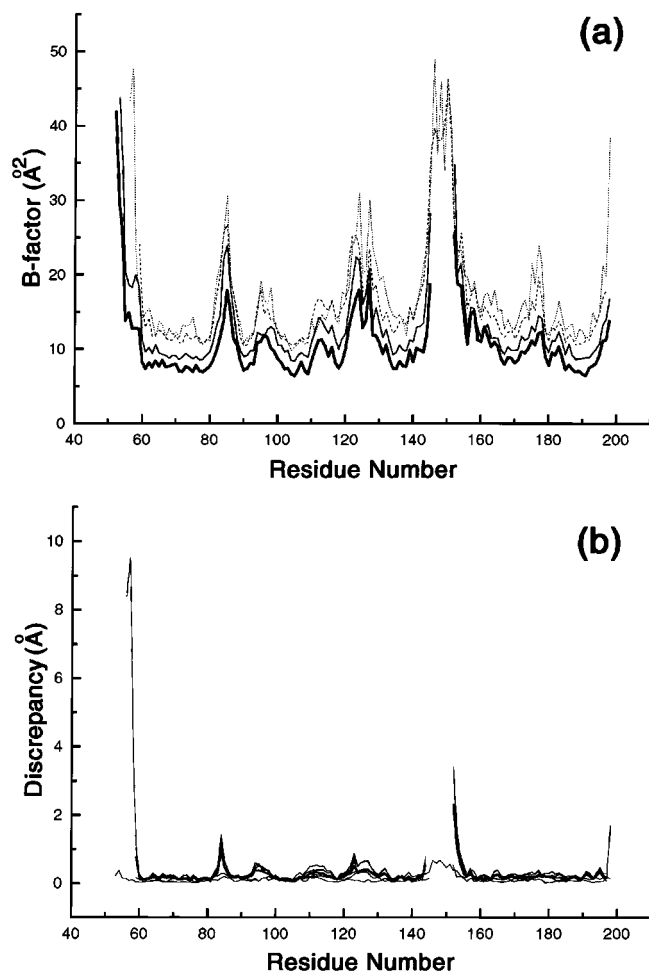


FIGURE 2: Thermal parameters and deviations between the ASV IN structures. (A) Distribution of the isotropic B factors in ASVIN-HEP (solid line), ASVIN-AS (thick solid line), ASVIN-CIT (dashed line), and D64N-CIT (dotted line). The plotted B factor values are those of the $C\alpha$ atoms. (B) The differences among the positions of equivalent $C\alpha$ atoms in different structures. Each of the four models described in this paper was aligned against the other structures (six pairs) on the basis of the $C\alpha$ atoms, with discrepancies among positions of equivalent $C\alpha$ atoms plotted as thin solid lines. For each $C\alpha$ position, the average discrepancy was calculated and the resulting distribution is plotted as a thick solid line. A comparison of the profiles shown in panels A and B clearly reveals a correlation between increasing B -factor values within a structure and greater $C\alpha$ position variation between the structures.

corresponding $C\alpha$ atoms is 0.54 \AA for Gly152, which is almost twice as large as the second largest deviation. Because Gly152 is adjacent to the disordered flexible loop, this difference may also reflect dynamic effects. For the side chains of 13 residues (Arg70, Val81, Ile88, Gln92, Val99, Ser124, Cys125, Met155, Glu157, Leu163, Asp165, Arg168, and Asn197) in ASVIN-AS and for the side chains of 11 residues (Arg70, Val81, Ile88, Gln92, Ser124, Cys125, Lys129, Leu163, Asp165, Glu187, and Asn197) in ASVIN-HEP, we modeled two discrete conformations. Besides Ile146 and Pro147, which are ordered in ASVIN-AS but disordered in ASVIN-HEP, the only other conformational differences were observed for the side chains of Gln59, Arg114, Lys129, and Arg161. Most of these residues are located on the surface of the molecule, and their side chains are characterized by B -factors that are higher than those for the adjacent residues.

We found somewhat larger differences between the two low-pH structures. One region with significant deviations

corresponds to residues Gly145–Ala154 (i.e., to the flexible active-site loop), with a maximum difference of 0.7 \AA in the $C\alpha$ atom positions. We found a similar deviation for the $C\alpha$ atoms of Gly123. One reason for these differences might be the lower resolution of ASVIN-CIT (1.42 \AA) compared with D64N-CIT (1.20 \AA), which led us to use a different thermal motion model in the refinement of each structure. Several residues (Val81, Ser150, Leu163, Asp165, and Met177 in ASVIN-CIT, and Arg70, Arg74, Val81, Arg114, Lys119, Lys129, Leu163, Asp165, Lys166, and Met177 in D64N-CIT) also have their side chains modeled in two discrete conformations. Other residues for which different conformations of the side chains are also observed are Arg95, Arg132, Ile146, and Lys178. Similar to the high-pH structures, all of these residues are located on the molecular surface. The substitution at position 64 in D64N-CIT might also contribute to the differences.

Even larger, yet consistent, differences are observed while comparing the structures corresponding to different pH values (Figures 1 and 3). Aside from the two most striking differences observed for the flexible loop region and the N-terminus (see above), there are additional regions of the IN molecule for which conformations seem to be pH controlled, the most important being the changes observed in and around the active-site area. These changes consist of rotation of the side chains of Lys119 (Figure 3A), Asp64, Asn122, Ile146, Pro147, and Gln153, as well as movement of Gly145 (Figure 3B). Differences found for Asp64 and Lys119 are discussed in detail below. Residues Gly145–Gln153 are part of the active-site loop, and the changes in their conformations are a direct consequence of the changes in the pH-controlled mobility of this loop. Another residue with a clear and reproducible conformational change is Asn122, although the reason for such variability is unknown. An additional, quite pronounced structural difference between the low-pH and high-pH structures is evident for the type I β -turn (34), Thr83–Ala84–Ser85–Ser86 (Figure 3C). Because this section is located near the N-terminus of the IN core domain, a change in the dynamic properties of the N-terminal residues can be a determining factor for this shift.

As described previously (14), ASV IN crystals grown in the presence of Hepes (the condition for both of the high-pH structures described here) bind this molecule in a specific fashion. Such binding triggers minor structural changes that are not directly pH dependent. Thus, smaller consistent differences (equivalent to $\sim 0.35\text{--}0.6 \text{ \AA}$ shifts of the $C\alpha$ atoms) are observed for the section His93–Thr97 (the loop linking the third β -strand and the first α -helix in the core domain of ASV IN). The Hepes molecule forms two hydrogen bonds ($\sim 3.10 \text{ \AA}$) with the side chain of Arg95. This side chain becomes much better ordered due to these interactions, and consequently, the whole His93–Thr97 loop is pulled slightly toward the ligand molecule. The changes observed for the side chain of Trp76 can also be attributed to the same factor.

Thermal Motion in the ASV IN Core Domain. The displacement parameters for all non-hydrogen atoms in the structures of ASVIN-HEP, ASVIN-AS, and D64N-CIT were refined anisotropically. The validity of this approach was confirmed in all cases by the associated decrease in free R -factor values during refinement. The average B factor for non-hydrogen protein atoms is directly proportional to the

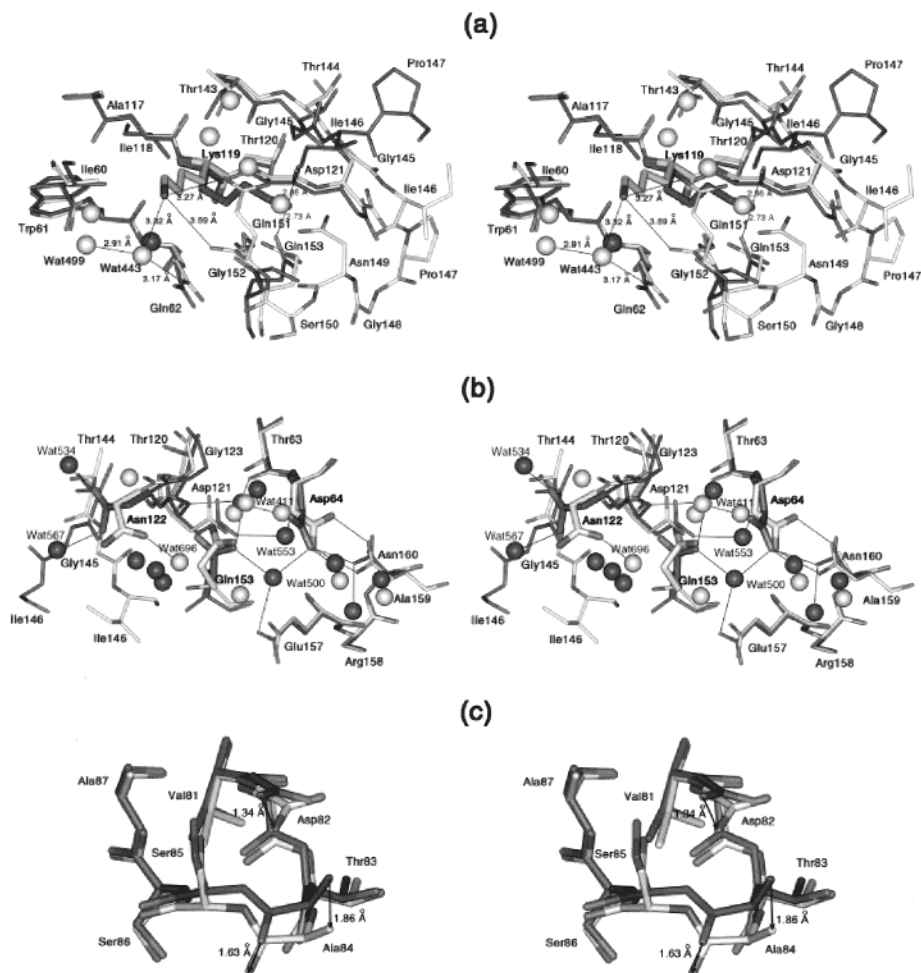


FIGURE 3: Stereo representation of the pH-dependent changes in the conformation of the residues located near the ASV IN active site. The C α atoms of the low-pH structure (D64N-CIT, dark gray) are superimposed on the high-pH structure (ASVIN-AS, light gray). (A) Two conformations of Lys119, shown as thicker lines. Hydrogen bonds observed for both conformations of Lys119 are also shown. The differences between the structures of the active site loops are very clear. (B) Different pH-controlled conformations of Asp64, Asn122, and Gln153, shown as thicker lines. For clarity, the lengths of the hydrogen bonds are not marked. (C) pH-triggered shift of the type I β -turn. The extent of this change is indicated by the three arrows.

highest experimental resolution. Thus, for ASVIN-AS, this is equal to 13.2 \AA^2 ; for ASVIN-HEP, it is 15.2 \AA^2 ; for D64N-CIT, it increases to 18.4 \AA^2 ; and finally it reaches a value of 19.9 \AA^2 for ASVIN-CIT. Although the absolute values of the B factors clearly contain experimental artifacts, their relative values within one model can be used to differentiate sections of the monomer according to their dynamic properties. These dynamic properties, represented as a graph of B factors for the C α atoms in all four models, are shown in Figure 2A. It is clear that, despite differences in data resolution and refinement protocols, all four plots have very similar profiles, indicating that the same sections of the molecule are most flexible. Aside from the N- and C-termini, there are three other regions within the ASV IN core with higher flexibility. One region is the active-site flexible loop (residues 144–154), the second is the section beginning with catalytic Asp121 and extending to the short helical fragment (the third helix in the ASV IN core domain), and the third region is the hairpin consisting of residues Thr83–Ala84–Ser85. These sections of the protein are all located around the active-site pocket (Figure 1A).

Conformation of the Active-Site Loop. Residues 144–154, forming the active-site flexible loop, have different dynamic properties under the various conditions that we studied. The

conformation of this section was unambiguously determined in two structures, ASVIN-CIT and D64N-CIT, and a similar electron density could also be seen for ASVIN-TRN. We found that in these structures, the active-site loop has a common conformation, different from any described previously (14, 20). As shown in Figure 4A, the electron density is very clear for nearly all residues in the loop. The average isotropic B factor for the non-hydrogen atoms of residues 144–152 is 38.8 \AA^2 in D64N-CIT, compared with 18.4 \AA^2 for all non-hydrogen atoms in the protein, indicating that this unambiguously placed segment is still more flexible than the rest of the protein. The loop residues form three well-defined secondary structure elements. The dipeptide Ala141–His142 creates a short β -strand that is part of the motif containing residues 117–119/124–138/141–142 (34). Residues Ile146–Asn149 form a type II β -turn. Beginning with Ser150, the next 24 residues form the longest helical fragment in the whole domain (the fourth α -helix). A number of hydrogen bonds are formed within the flexible loop. Most of them are formed by main-chain atoms: Ile146(O)–Asn149(N) (3.06 \AA), Asn149(O)–Gln153(N) (2.91 \AA), Ser150(O)–Ala154(N) (2.99 \AA), and Ser150(O)–Gln153(N) (3.48 \AA). An additional hydrogen-bonded interaction is formed by the side-chain oxygen of Asn149 and Gly152(N)

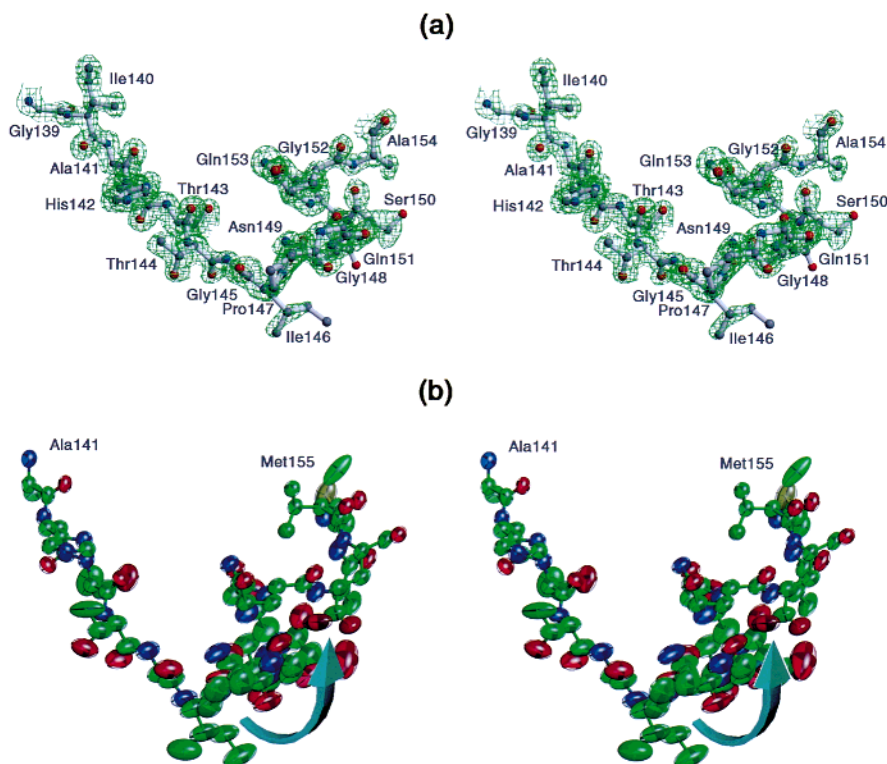


FIGURE 4: Active-site loop in the structure ASVIN-CIT. (A) $2F_o - F_c$ electron density for the final model, contoured at the 1σ level. This panel, as well as Figure 5A, were prepared with the program BOBSCRIPT (37) and rendered with RASTER3D (38). (B) Thermal ellipsoids plotted on a 20% probability level using the program ORTEX (39).

(2.91 Å). All of these hydrogen bonds are characterized by very good stereochemistry. Other stabilizing interactions are present between the loop and the rest of the protein. The most stabilizing interactions appear to be a number of stereochemically well-defined hydrogen bonds: Gly145(O)—Asn122(N) (2.92 Å), Gln151(O)—Lys119(NZ) (3.11 Å), Gln151(O)—Gln62(NE2) (3.37 Å), and Gln153(NE2)—Thr63(O) (3.31 Å).

Because the anisotropic displacement parameters approximate the degree of atomic disorder and its dominant direction, we generated thermal ellipsoids for the flexible loop, which are shown in Figure 4B. The concerted movement of the atoms clearly suggests the motion of the whole loop.

Structures of ASVIN-TRN and D64N-TRN. To extend our understanding of the pH-induced structural changes within the catalytic domain of ASV IN, we collected atomic resolution X-ray data for crystals of ASVIN-TRN and D64N-TRN that were transferred to solutions with new conditions (see Experimental Procedures and Table 1). Crystals of ASVIN-TRN were transferred to an acetate buffer-based solution containing 50 mM $MnCl_2$ with the same pH. Analysis of the electron density maps calculated for the partially refined structure of ASVIN-TRN clearly revealed the location of the flexible loop residues (144–152) that were omitted from the refinement, with the conformation of this segment being virtually identical to that observed for ASVIN-CIT. The most important difference was seen within the active site of ASVIN-TRN, where we could easily identify a peak corresponding to the Mn^{2+} cation bound to the side chain of Asp64, the latter present in the active conformation that was identical to that found in ASVIN-HEP. Comparison of the $2F_o - F_c$ electron density map of ASVIN-TRN with

the model of ASVIN-CIT did not show other significant structural changes. In contrast, the side chain of Asn64 had an orientation identical to that observed in D64N-CIT. However, in D64N-TRN, residues 144–152 did not have an interpretable electron density, even though the X-ray data for this crystal were of better quality and higher resolution than for D64N-CIT, whose active-site loop had a clearly defined conformation. Further comparisons indicated a number of additional differences between D64N-TRN and D64N-CIT, which resulted from the pH change accompanying the crystal transfer and which were similar to those observed in the comparison of ASVIN-CIT and ASVIN-HEP.

C—H \cdots O Hydrogen Bonds. Careful analysis of the models revealed a set of unusual hydrogen bonds that are very seldom reported for proteins. We found that His142 forms four hydrogen bonds, utilizing four atoms (N δ 1, N ϵ 2, C δ 2, and C ϵ 1) of its side-chain ring (see Figure 5). Thus, in addition to the expected N—H \cdots O hydrogen bonds, we observed the rarer C—H \cdots O hydrogen bonds. The high resolution of the data and the consistency among the different structures validate this result. His142 and the surrounding residues that participate in this unusual hydrogen bonding are structurally very conserved in the four models presented here.

Additional Binding Sites in ASV IN. Analysis of the electron density for all of the structures revealed a few large peaks that could not be described as either amino acid residues or water. We surmised the identity of the molecules occupying the appropriate sites from the shape and size of the electron density, the interatomic contacts with neighboring protein atoms, and the contents of the crystallization medium. Hepes molecules were previously reported in the

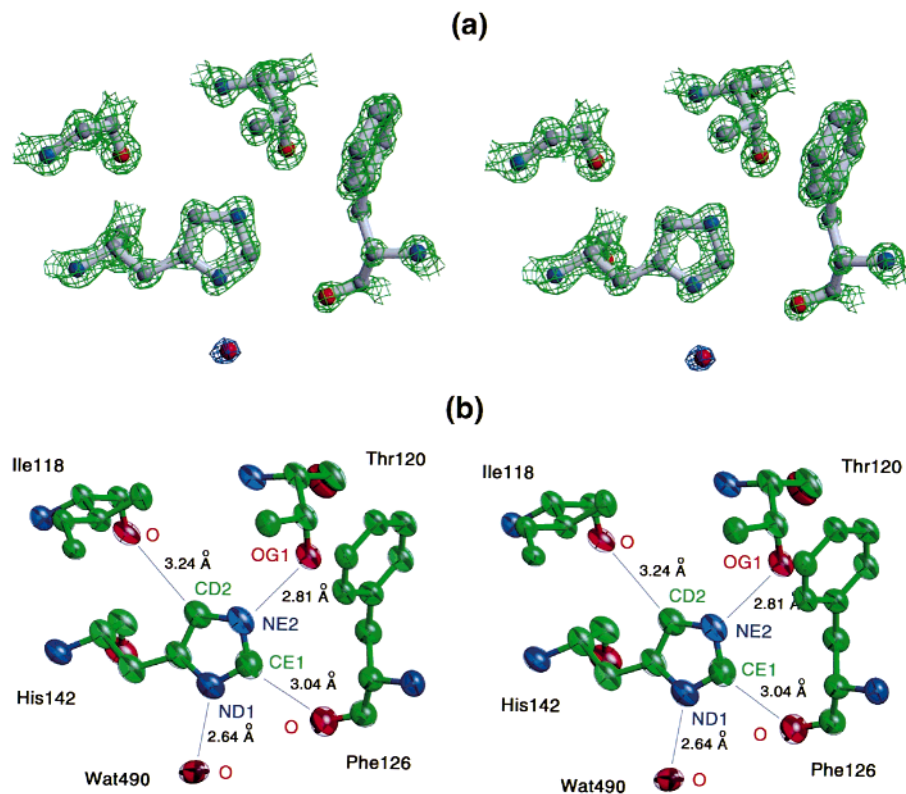


FIGURE 5: Hydrogen-bonding patterns of His142 in ASVIN-AS. (A) Electron density contoured at the 1.3σ level (green) for all amino acids, whereas the blue density covering Wat490 is contoured at the 0.65σ level. This mode of presentation was chosen because the water site is half-occupied. Map resolution 1.02 Å. (B) Thermal ellipsoids plotted on a 20% probability level. The hydrogen bond interactions are marked as gray lines.

crystals containing this buffer (14). Another molecule identified in this study is the citrate anion, located at the interface between two protein monomers and distant from the active site. This molecule is stabilized by several hydrogen bonds with protein and water molecules, as well as by salt bridges between its carboxylate groups and arginines 132 and 179.

We interpreted another feature of the electron density as a glycerol molecule located close to the crystallographic 2-fold axis. Glycerol interacts with the protein and solvent as well as with a symmetry-related glycerol molecule via hydrogen bonds. Interactions with the enzyme are through hydrogen bonds between Val 90(O) and a hydroxyl group of glycerol and by hydrophobic contacts between the carbon atoms of the ligand and the side chains of Val89, Thr91, Thr107, and Met193. As with the citrate anion, the glycerol molecule is bound away from the enzyme's active site.

Solvent Structure. About 200 water molecules were identified independently in each of the four structures on the basis of their electron densities (see Table 1). The average B factors for the solvent vary between 20.0 \AA^2 for ASVIN-HEP and 36.8 \AA^2 for ASVIN-CIT, indicating some correlation with the resolution of the X-ray data. Water molecules conserved among the four refined structures were located after superimposing these structures on the $C\alpha$ coordinates and were assumed to occupy a common position if four water molecules, each from a different structure, were located within a 0.6 \AA radius sphere. We performed the calculations with a computer program written for this purpose (J. L., unpublished material), and we confirmed their correctness by subsequent visual inspection. This procedure identified

82 structurally conserved water molecules. The average B factors for these subsets are 19.4 \AA^2 (ASVIN-AS), 19.0 \AA^2 (ASVIN-HEP), 27.9 \AA^2 (ASVIN-CIT), and 22.9 \AA^2 (D64N-CIT) and are, on average, lower by $\sim 20\%$ than the corresponding values shown in Table 1 for all solvents. A detailed inspection revealed that 21 of these conserved water molecules participate in four near-tetrahedrally arranged hydrogen bonds, and for six water molecules, we could identify four hydrogen bonds with somewhat less ideal geometry. An additional 31 water molecules participate in three hydrogen bonds, and 22 water molecules are incorporated into the structures via two hydrogen bonds. Only two conserved water molecules interact with the protein via a single hydrogen bond. About one-fourth of these conserved water molecules (21) are located within the crystallographic dimer interface, and four of them (Wat402, Wat413, Wat470, and Wat472) clearly play a direct role in stabilizing dimerization.

Using the same criteria, we identified 29 solvent sites common to three out of four of the refined structures, and an additional 76 water molecules conserved between pairs of structures. It is likely that, despite the very high resolution and good quality of all the structures presented here, some conserved water molecules were not accounted for, due in part to the fact that all the structures contain disordered regions and to the limits of the experimental information achieved in this study.

DISCUSSION

The results presented here clearly show that the dynamic properties of the active-site loop change dramatically as a

function of pH within the range of 6.0–7.5. Because this pH range coincides with the known pH vs activity assay profile for IN (22), there might be a direct relationship between the pH-dependent changes in the conformation of the active-site loop and the enzymatic activity of IN. As previously described (22), a change of protonation followed by a conformational change is observed in the core domain of native ASV IN for the Asp64 side chain around pH 6.2. Thus, we can see that pH-triggered structural changes of this protein are quite extensive, especially around its active site (Figure 3). Studies of enzymatic activity as a function of pH (22) indicated that the catalytic activity of ASV IN decreases dramatically at pH below 6.0. Therefore, it is very likely that some or all of the structural changes observed in the region of the active site are responsible for the changes in the enzymatic activity.

The conformation of the active-site loop described here for D64N-CIT and ASVIN-CIT is different from any that was previously described (14, 20), showing three major new features. First, this loop is folded in a compact fashion on the surface of the molecule (Figure 1). All residues have conformations within the favorable regions of the Ramachandran plot (35), in contrast to Ser150 in previous models, which was placed in an unfavorable region. Finally, the average temperature factor for the main-chain atoms within this loop is 36.0 Å² (or 38.8 Å² for all atoms of residues 144–152), compared with 60 Å², the lowest previously reported (14) (for the structure with PDB code 1asw, also solved with a frozen crystal). Because the crystallization conditions that we used in the present studies were the same as those reported in the past, these discrepancies require further explanation. We analyzed the data on the conformations of the active-site loop, including all structures generated during our previous experiments, and compared them with the atomic resolution structures presented here. In the initial papers describing the structure of the ASV IN core, two models for the active site loop were proposed: one for crystals grown in the presence of ammonium sulfate as the precipitant, and the other for PEG4000 as the precipitant (14, 20). No other factors, such as pH or the presence of metal cations, were reported to alter loop conformation. It is important to note that none of the previous structures of the ASV IN core were refined at a resolution higher than 1.7 Å. We reanalyzed the diffraction data, refinement protocols, and electron density maps for several structures that included coordinates for the flexible loop. We found that all three elements were very consistent with those described here, except for the discrepancies resulting from the very significant difference in the resolution limit of the X-ray data. As a result of this analysis, we could not confirm the previously published loop conformations. Therefore, on the basis of the atomic resolution structures of the core domain of ASV IN, we unambiguously identified only one novel conformation of the active-site loop. This conformation is present at pH lower than 6.0–6.5 and is usually accompanied by conformational changes of several side chains, including those of Asp64 (Figure 3B) and Lys119 (Figure 3A). At pH higher than 7.0–7.5, the active-site loop becomes disordered beyond the limits of detection by X-ray crystallography, and this property is very likely essential for enzyme functionality.

We conducted two additional experiments to better understand the possible correlation between the conforma-

tions of the active-site loop and the conformations of other residues of the core domain. The high-quality electron density maps of ASVIN-TRN and D64N-TRN were most informative, although the refinement was not completed. It is clear that the protonation state of Asp64 is not correlated with any specific conformation of the active-site loop and that Asp64 does not participate in stabilizing the low-pH conformation of residues 144–152. These experiments also confirmed our previous observation that divalent metal cations (Mn²⁺, Mg²⁺) can stabilize the active conformation of Asp64 at pH as low as 5.5 (J.L., unpublished data).

In the second experiment, we additionally show that the active-site loop becomes nearly completely disordered upon an increase of pH, despite the fact that the side chain of Asn64 in D64N-TRN remains in its inactive conformation, incapable of binding divalent cations. Even though both experiments were associated with very high-resolution X-ray data (see Table 1), in neither case could we detect the presence of stable conformations of the flexible active-site loop that were different from those described here.

We analyzed in detail the interactions responsible for the stabilization of the low-pH conformation of the active-site loop. Because Asp64 was shown not to directly stabilize the loop, we tried to identify other residues that also might change their protonation state around pH 6.5 and be responsible for changing the dynamic properties of this loop. Only one residue near the structurally variable fragment 144–152 (His142) could potentially change its protonation state in the pH range of interest. However, His142 does not interact directly with any of the flexible loop residues, and thus, it is unlikely that its protonation determines the conformation of the loop. We extended a similar analysis over the residues that contribute their side chains to formation of hydrogen bonds with the active-site loop. Three charged residues (Lys119, Asp121, and Glu157) are evident in the vicinity of fragment 144–152. The two acidic residues are part of the IN active site itself. Similar to Asp64, their protonation probably affects the enzyme's affinity for binding the divalent metal cations that are necessary for activity. Mutation studies of these residues are currently under way and will address this problem in detail. For the current analysis, it is important to mention that only the side chain of Glu157 is located within hydrogen-bonding distance of the loop, ~3.5 Å from Gly148(N). Even in this case, however, the potential hydrogen bond would be questionable due to the unfavorable orientation of the contributing atoms. The side chain of Asp121 points in a direction opposite to that of the flexible loop. Moreover, the conformations of the side chains of both Asp121 and Glu157 are not affected by the pH changes, as seen when comparing the structures of ASVIN-HEP and ASVIN-CIT.

We found a very different situation, however, for Lys119. We could clearly identify a conformational change for the side chain of this residue as a function of pH. In one conformation (pH < 6.5), Lys119(N ξ) creates a stereochemically favorable hydrogen bond with Gln151(O), which we observed for both D64N-CIT (N–H \cdots O, 3.11 Å) and ASVIN-CIT (N–H \cdots O, 3.53 Å) and a second interaction with Gln151(OE1) (N–H \cdots O, 3.27 Å), detected only in ASVIN-CIT. These interactions clearly stabilize the conformation of the flexible loop observed at low pH. Such stabilization is not provided by the side chain of Lys119 in

the alternate conformation, where the N ξ atom is located at a distance of 2.93 Å toward Ala120(O), a residue in a structurally conserved region. Currently, it is not possible to prove that the interaction with Lys119 is the sufficient and only factor for stabilizing of the conformation observed for residues 144–152. Additionally, we do not know whether there is any relationship between the two conformations of Lys119 and the protonation state of its side chain. Finally, the available results are insufficient to remove all ambiguity about a direct link between specific structural changes resulting from the pH change and the activity of the enzyme. This analysis, however, suggests a possible function for Lys119 for the activity of the enzyme, making this residue a very interesting candidate for mutational studies.

In HIV-1 IN, the position equivalent to ASV IN Lys119 is occupied by His114. Although histidine is also basic, its geometry is significantly different from that of lysine. Because the side chain of Lys119 is extended when it stabilizes the low pH loop conformation, a similar interaction cannot be present in HIV-1 IN without significant conformational changes of either the active-site loop, the fragment containing His114, or both. The active-site loop in HIV-1 IN has been observed in two structures (PDB codes 1bis and 2itg), describing two different crystal forms of the catalytic core of this enzyme. Although in both cases the conformations of the loop are very different, in neither one does His114 interact similarly to Lys119 in ASV IN. The loop conformations reported for the HIV-1 IN structures are also very different from the one described here for ASV IN. Because the primary structure of this segment of the catalytic domain is highly conserved in both enzymes, we could ask whether any of the reported loop conformations are biologically relevant. A conclusive answer requires new information, e.g., the structure of the complex between IN and its substrate or possibly the structure of the entire enzyme.

While analyzing the anisotropic displacement parameters for the loop atoms, we observed that all residues within the most flexible region of the loop move in a concerted fashion (Figure 4B) and the direction of the loop movement is approximately perpendicular to the active-site pocket. Two hypotheses can be proposed from these observations. First, a concerted movement of the loop residues suggests that this motif has a well-defined conformation, i.e., it is stabilized by interactions within itself. Additionally, this observation may correspond to the movement of residues associated with pH changes, possibly similar to the loop dynamics present at some stages of DNA substrate binding (or product release) at the active site of the enzyme.

Although hydrogen bonds with carbon atoms as donors (C–H \cdots O) are not very rare in the structures of small molecules, especially nucleotides and nucleosides, they are very seldom reported for proteins (36). In the structures presented here, all four hydrogen bonds formed by the ring of His142 are very apparent. We can rule out the possibility that these interactions represent a mixture of two rotamers of His142, on the basis of the behavior of the anisotropically refined temperature factors of this residue. The results are in agreement with the data of Jeffrey and Saenger (36), who reported that for C–H \cdots O hydrogen bonds the distance between H and O is almost always larger than 2.0 Å (usually ~2.3 Å), whereas it is shorter for N–H \cdots O hydrogen bonds (mostly in the range 1.55–1.90 Å). Thermal ellipsoids

(Figure 5B) indicate that the vibrations for all four acceptor atoms [Ile118(O), Thr120(O γ 1), Phe126(O), and Wat490(O)] have a relatively isotropic distribution. This observation is not compatible with a postulate of 2-fold disorder for the hydrogen donors. However, the possible significance of the extensive interactions of His142 and the surrounding residues is not clear. One appealing possibility is that His142 plays a key role in stabilizing the local structure of ASV IN: it interacts with three residues (Ile118, Thr120, and Phe126) located in the direct vicinity of the enzyme's active site.

ACKNOWLEDGMENT

We thank Anne Arthur for editorial assistance.

REFERENCES

- Katz, R. A., and Skalka, A. M. (1994) *Annu. Rev. Biochem.* 63, 133–173.
- Prasad, G. S., Stura, E. A., McRee, D. E., Laco, G. S., Hasselkus-Light, C., Elder, J. H., and Stout, C. D. (1996) *Protein Sci.* 5, 2429–2437.
- Wagaman, P. C., Hasselkus-Light, C. S., Henson, M., Lerner, D. L., Philips, T. R., and Elder, J. H. (1993) *Virology* 196, 451–457.
- Dauter, Z., Persson, R., Rosengren, A. M., Nyman, P. O., Wilson, K. S., and Cedergren-Zepezaue, E. S. (1999) *J. Mol. Biol.* 285, 655–673.
- Coleman, R. L., and Holtzer, C. (1998) HIV-Related Drug Information (Section 4). In *The AIDS Knowledge Database* (Cohen, P. T., Sande, M. A., and Volberding, P. A., Eds.) University of California, San Francisco General Hospital, <http://hivinsite.ucsf.edu/akb/1997/>.
- Appelt, K., Bacquet, R. J., Bartlett, C. A., Booth, C. L., Freer, S. T., Fuhry, M. A., Gehring, M. R., Herrmann, S. M., Howland, E. F., and Janson, C. A. (1991) *J. Med. Chem.* 34, 1925–1934.
- Vink, C., and Plasterk, R. H. (1993) *Trends Genet.* 9, 433–438.
- Goff, S. P. (1992) *Annu. Rev. Genet.* 26, 527–544.
- Cai, M., Zheng, R., Caffrey, M., Craigie, R., Clore, G. M., and Gronenborn, A. M. (1997) *Nat. Struct. Biol.* 4, 567–577.
- Eijkelenboom, A. P., van den Ent, F. M., Vos, A., Doreleijers, J. F., Hård, K., Tullius, T. D., Plasterk, R. H., Kaptein, R., and Boelens, R. (1997) *Curr. Biol.* 7, 739–746.
- Lodi, P. J., Ernst, J., Kuszewski, J., Hickman, A. B., Engelman, A., Craigie, R., Clore, G. M., and Gronenborn, A. M. (1995) *Biochemistry* 34, 9826–9833.
- Eijkelenboom, A. P., Lutzke, R. A., Boelens, R., Plasterk, R. H., Kaptein, R., and Hård, K. (1995) *Nat. Struct. Biol.* 2, 807–810.
- Dyda, F., Hickman, A. B., Jenkins, T. M., Engelman, A., Craigie, R., and Davies, D. R. (1994) *Science* 266, 1981–1986.
- Bujacz, G., Jaskólski, M., Alexandratos, J., Wlodawer, A., Merkel, G., Katz, R. A., and Skalka, A. M. (1995) *J. Mol. Biol.* 253, 333–346.
- Bujacz, G., Alexandratos, J., Wlodawer, A., Merkel, G., Andrade, M., Katz, R. A., and Skalka, A. M. (1997) *J. Biol. Chem.* 272, 18161–18168.
- Grindley, N. D. F., and Leschziner, A. E. (1995) *Cell* 83, 1063–1066.
- Maignan, S., Guilloteau, J. P., Zhou-Liu, Q., Clement-Mella, C., and Mikol, V. (1998) *J. Mol. Biol.* 282, 359–368.
- Goldgur, Y., Dyda, F., Hickman, A. B., Jenkins, T. M., Craigie, R., and Davies, D. R. (1998) *Proc. Natl. Acad. Sci. U.S.A.* 95, 9150–9154.
- Heuer, T. S., and Brown, P. O. (1997) *Biochemistry* 36, 10655–10665.
- Bujacz, G., Jaskólski, M., Alexandratos, J., Wlodawer, A., Merkel, G., Katz, R. A., and Skalka, A. M. (1996) *Structure* 4, 89–96.

21. Lubkowski, J., Yang, F., Alexandratos, J., Wlodawer, A., Zhao, H., Burke, T. R., Jr., Neamati, N., Pommier, Y., Merkel, G., and Skalka, A. M. (1998) *Proc. Natl. Acad. Sci. U.S.A.* 95, 4831–4836.
22. Lubkowski, J., Yang, F., Alexandratos, J., Merkel, G., Katz, R. A., Gravuer, K., Skalka, A. M., and Wlodawer, A. (1998) *J. Biol. Chem.* 273, 32685–32689.
23. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol.* 276, 307–326.
24. Brünger, A. T. (1992) *Nature* 355, 472–474.
25. Brünger, A. (1992) *X-PLOR Version 3.1: A System for X-ray Crystallography and NMR*, Yale University Press, New Haven.
26. Jones, T. A., and Kieldgaard, M. (1997) *Methods Enzymol.* 277, 173–208.
27. Sheldrick, G. M., and Schneider, T. R. (1997) *Methods Enzymol.* 277, 319–343.
28. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr.* 26, 283–291.
29. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) *Acta Crystallogr., Sect. D* 53, 240–255.
30. Lamzin, V. S., and Wilson, K. S. (1997) *Methods Enzymol.* 277, 269–305.
31. Engh, R., and Huber, R. (1991) *Acta Crystallogr., Sect. A* 47, 392–400.
32. Sevcik, J., Dauter, Z., Lamzin, V. S., and Wilson, K. S. (1996) *Acta Crystallogr., Sect. D* 52, 327–344.
33. Cohen, G. E. (1997) *J. Appl. Crystallogr.* 30, 1160–1161.
34. Hutchinson, E. G., and Thornton, J. M. (1996) *Protein Sci.* 5, 212–220.
35. Ramakrishnan, C., and Ramachandran, G. N. (1965) *Biophys. J.* 5, 909–933.
36. Jeffrey, G. A., and Saenger, W. (1994) in *Hydrogen Bonding in Biological Structures* (Jeffrey, G. A., and Saenger, W., Eds.) Springer-Verlag, Berlin, Heidelberg, Germany.
37. Esnouf, R. M. (1997) *J. Mol. Graphics* 15, 133–138.
38. Meritt, E. A., and Murphy, M. E. P. (1994) *Acta Crystallogr., Sect. D* 50, 869–873.
39. McArdle, P. (1995) *J. Appl. Crystallogr.* 28, 65.

BI991362Q