Structural models of ribonuclease H domains in reverse transcriptases from retroviruses

Haruki Nakamura, Katsuo Katayanagi, Kosuke Morikawa and Morio Ikehara
Protein Engineering Research Institute, 6-2-3 Furuedai, Suita, Osaka 565, Japan

Received January 7, 1991; Revised and Accepted February 26, 1991

ABSTRACT

Tertiary models of ribonuclease H (RNase H) domains in reverse transcriptases (RTs) from Moloney murine leukemia virus (MuLV) and human immunodeficiency virus (HIV-1) were built based upon the X-ray structure of RNase H from *Eschericia coli* (*E. coli* RNase H). In two models of RT-RNase H domains, not only active site residues but also residues, which construct a hydrophobic core and hydrogen bonds, are located in the same positions as those of *E. coli* RNase H. The whole backbone structure and the electrostatic molecular surface of MuLV RT-RNase H model are similar to those of *E. coli* RNase H. On the contrary, HIV-1 RT-RNase H model lacks the third helix and the following loop, resulting no positive charge clusters around the hybrid recognition site. Referring the complex models of RTs with their substrate hybrid, the interaction between DNA-polymerase and RNase H domains in RTs was discussed.

INTRODUCTION

A reverse transcriptase (RT) from retroviruses has an essential role in their life cycle, copying their RNA genome and producing a double stranded DNA, which is then incorporated into the double stranded DNA of an infected cell [1]. It is now commonly understood that RTs from retroviruses are composed of two functional domains; One is RNA-dependent DNA-polymerase and the other is ribonuclease H (RNase H) [2].

Oyama *et al.* propose a bifunctional model of RTs from analyses of cleavage modes for a synthetic mRNA for ovalbumin and complementary oligodeoxynucleotides [3]. They conclude that RNase H active sites of RTs are located close to 3'-terminus of RNA strand of the substrate hybrid about 10 base pairs downstream from the polymerase active site. Similar finding was made for RT from human immunodeficiency virus (HIV-1), too [4]. However, since any overall tertiary structure of RT has not yet been revealed, precise discussion cannot be made about the structure-function relationships of RTs.

Doolittle *et al.* [5] have found remarkable resemblances in amino acid sequences between RNase H from *Eschericia coli* (*E. coli* RNase H) and the RNase H domain of RTs from many retroviruses including Moloney murine leukemia virus (MuLV) and HIV-1. Recently, the tertiary structure of *E. coli* RNase H was revealed by two independent groups of Katayanagi *et al.* [6] and Yang *et al.* [7] Moreover, the active site of *E. coli* RNase H has been identified by protein engineering studies [8], and the substrate recognition mode of the enzyme has been studied by the present authors [9].

Although only weak homologies are found in the whole amino acid sequences between RNase H domains in RTs from retroviruses and *E. coli* RNase H, the essential amino acids for the enzymatic activity in *E. coli* RNase H are well conserved in all RTs [5]. Assuming similar three dimensional structures among those enzymes, it is now possible to construct meaningful models of those RNase H domains in retroviral RTs, starting from the structure of *E. coli* RNase H.

In the present paper, MuLV and HLV-1 RT-RNase H domains were modeled, based upon the current structural and functional information about *E. coli* RNase H. The interaction between RNase H and DNA-polymerase domains is also discussed from the model structures.

METHODS

Starting from the Doolittle alignment [5] of amino acid sequences among RNase H domains of RTs from retroviruses and *E. coli* RNase H, the alignments were rearranged. It is generally difficult to make a definite alignment when the homology between two amino acid sequences is less than 30%. Since the tertiary structure of *E. coli* RNase H has been solved, it is meaningful to realign the sequence so that secondary and tertiary structure should not be changed much. Any large gaps were not expected in the middle of secondary structures, and inner charged residues were avoided. In Fig. 1, the rearranged alignment is shown. The ratios of the conserved amino acids in the rearranged alignment for MuLV and HIV-1 RT-RNase H domains are 26% and 29%, respectively. In the previous multiple alignment by Doolittle *et al.* [5], they were 21% and 20%, respectively. As shown in Fig. 1, it is possible to align the amino acid sequence of MuLV RT-RNase H with that of *E. coli* RNase H without significant deletions or insertions. On the contrary, in the sequence of HIV-1 RT-RNase H, there is a large deletion around the associated third helix (αIII) of *E. coli* RNase H.

* To whom correspondence should be addressed

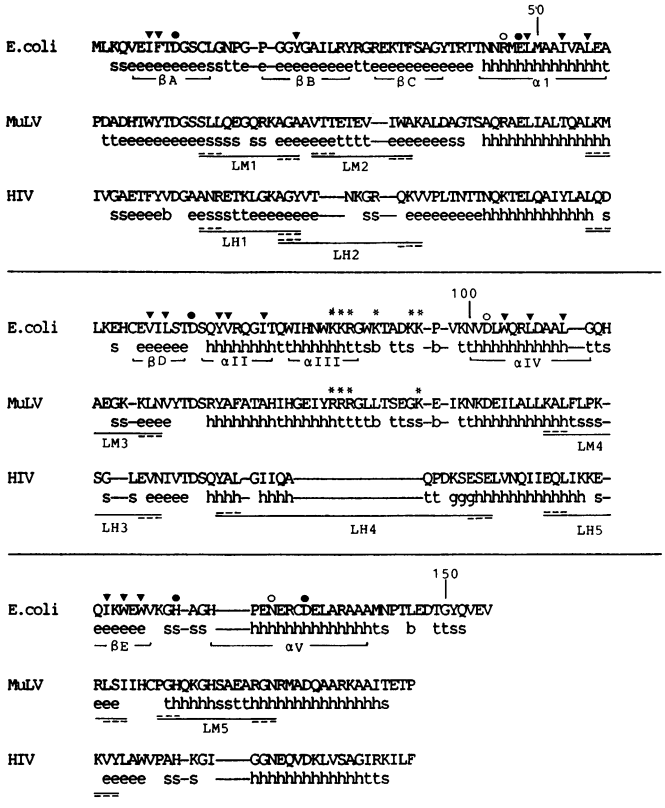

Figure 1. Alignment of sequences of *E. coli* RNase H and RNase H domains of RTs from MuLV and HIV-1. Filled circles indicate well conserved residues in most RTs, which are crucial for the activity of *E. coli* RNase H. Open circles indicate conserved residues, which stabilize the tertiary structure of *E. coli* RNase H by hydrogen bonds and salt bridges. Filled triangles are conserved hydrophobic and aromatic groups, constructing a hydrophobic core. Stars are characteristic basic amino acids, generating positive potentials around the minor domain (αIII and its following loop). Lowercase letters indicate conformations/secondary structures (h, α-helix; g, 3-to-10-helix; e, β-strand; b, β-bridge; t, 3-, 4- or 5-turn; s, bend; space, straight piece not in β-structure) calculated by a program DSSP developed by Kabsch and Sander [24]. For *E. coli* RNase H, these structures are from the X-ray structure solved by Katayanagi *et al.* [6] The secondary structures defined by them are also indicated under the lowercase letters. For MuLV and HIV-1 RT-RNase H, secondary structures are from the current models. Local loops from LM1 to LH5 for MuLV and HIV-1 RT-RNase H models are indicated by bars. Pre- and post-loops are indicated by bars with upper and lower dotted lines.

Applying the loop search method proposed by Jones and Thirup [10] amino acid backbone structures in the regions around such deletions and insertions were constructed. These local regions hereafter called 'loops' are also indicated in Fig. 1. By our house made program; FRGMNT, suitable structures for each loop were picked up among lots of possible loops in protein structures stored in Protein Data Bank (PDB) [11]. FRGMNT evaluated the root mean square deviations (r.m.s.d.) of backbone structures of the six residues on the both sides (three pre-loop residues and three post-loop residues) surrounding each loop, between *E. coli* RNase H and the superposed loop of protein structures in PDB. The most suitable loops were selected, so that they have no bad contacts with other parts of the model enzymes and the secondary structures of their pre- and post-loops are the same as those in *E. coli* RNase H. After the backbone structures were fixed, side chains were generated in order that their directions were similar to those in the original structure of *E. coli* RNase H.

Then, the initial three dimensional structures were optimized by minimizing their conformation energies using our house made program; PRESTO. Each optimization procedure was as follows. At first, 2000 steps of steepest descent minimization was carried out, fixing all backbone structures of the framework, which is defined as all the regions except loops indicated in Fig. 1. In this calculation, the repulsive forces proportional to squares of the distances between each pairs were used instead of Lennard-Jones or Coulombic interaction forces. Second, further 5000 steps of steepest descent minimization was carried out, fixing the same atoms as those in the above minimization. The Lennard-Jones and Coulombic interaction forces of AMBER force field [12] were applied. The electrostatic potential was calculated with a distance dependent dielectric constant ($\epsilon = 2.0 \times d_{ij}$; d_{ij} is a distance between two non-bonded atom pair). These forces were used after this stage. By this calculation, most of bad contacts were repaired. Third, further 5000 steps of conjugate gradient minimization was carried out, without any fixed atoms but with distance restraints among side chain atoms, the residues of which are conserved between *E. coli* RNase H and the RT-RNase H. The restraints were for residues, at least one atom pair in which are located nearer than 5.0 A. Finally, until the root mean square force became less than 0.2 kcal/mol/A, conjugate gradient minimization was carried out without any fixed atoms or distance restraints.

After the model structures were built, electrostatic molecular surfaces of the enzymes were calculated by solving Poisson-Boltzmann equations following the procedure of Nakamura and Nishida [13]. The results were displayed on the computer graphics screen (E & S PS390) as the dot surfaces of the enzymes using programs MS [14] and FRODO [10].

All the minimizations and electrostatic potential calculations were performed by FACOM VP400-E (Fujitsu, Co., Ltd.). The surface dots were generated on VAX/VMS-8810 (DEC).

RESULTS

Table 1 shows the most suitable loop structures found from PDB structural database for local sequences indicated in Fig. 1. For short loops, it is possible to find out suitable loop structures, which have small r.m.s.d. of the pre- and post-loops in *E. coli* RNase H.

Figures 2 -a) and -b) show the backbone structures of the current tertiary models of MuLV and HIV-1 RT-RNase H domains superposed on that of *E. coli* RNase H, respectively. The framework structures deviated a little in the optimization procedure. The r.m.s.d. between the coordinates of the backbone atoms in the framework of *E. coli* RNase H and those of MuLV and HIV-1 RT-RNase H domains are both 1.4 A.

Several amino acids significantly conserved in RT-RNase H domains are also shown in Figs. 2 -a) and -b). In *E. coli* RNase H, they are Asp 10, Glu 48, Arg 46, Asp 70, Asp 102, His 124, Asn 130 and Asp 134. These residues are well conserved from a structural view point, too. The secondary structures of these models are shown in Fig. 1, following the definition by Kabsch and Sander [24].

The whole backbone structure of MuLV RT-RNase H model is very similar to that of *E. coli* RNase H. On the contrary, the model of HIV-1 RT-RNase H domain lacks a 'minor domain' called by Katayanagi *et al.* [6] around the third helix (αIII) and the following loop before the fourth helix (αIV). It should also be noted that the lengths of the second and the third β-strands

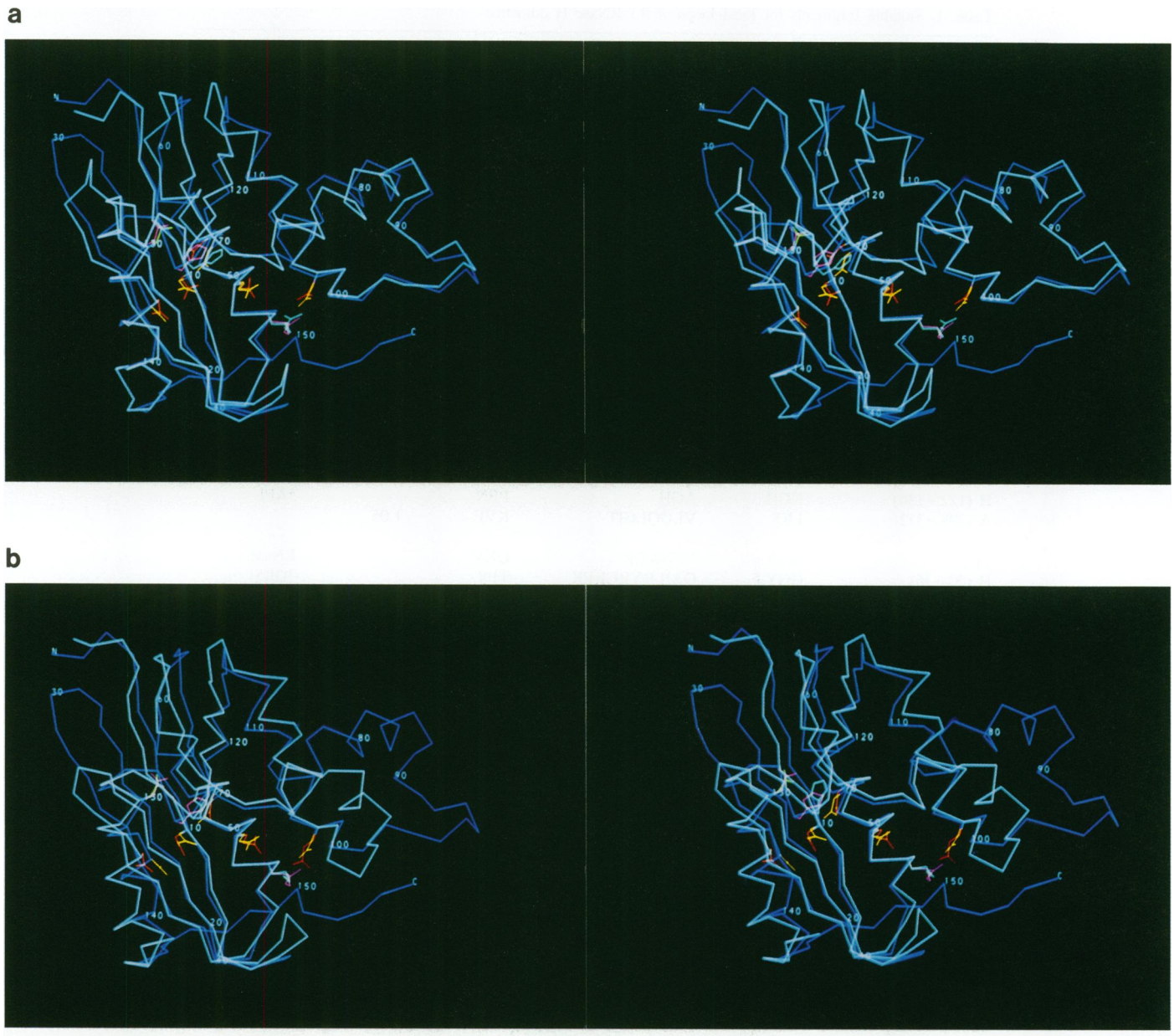

Figure 2. Stereo drawings of the backbone Cα structures of the current models of (a) MuLV and (b) HIV-1 RT-RNase H domain (white lines) superposed on *E. coli* RNase H (blue lines) [6]. Side chains of Asp 10, Glu 48, Asp 70, Asp 102 and Asp 134 in *E. coli* are shown by red lines and those of Arg 46, His 124 and Asn 130 are green lines. Side chains of the associated residues in (a) MuLV and (b) HIV-1 RT-RNase H are shown by yellow and purple lines. Numbers in figures are amino acid residue numbers in *E. coli* RNase H.

(βB and βC) are shorter in both the two models of RT-RNase H domains than those of *E. coli* RNase H.

In Fig. 3 -a), an electrostatic molecular surface of *E. coli* RNase H is illustrated by colored dots. Those surfaces of MuLV and HIV-1 RT-RNase H models are also shown in Figs. 3 -b) and -c), respectively.. Here, blue color indicates positive potential and red color negative. All electrostatic potentials were calculated without considering any contributions from the bound Mg^{2+} ion [6]. The features of potential distribution of MuLV RT-RNase H are similar to those of *E. coli* RNase H. Especially, they are remarkably positive around the minor domain at the upper right of Figs. 3 -a) and -b). On the contrary, in case of HIV-1 RT-RNase H, no significantly positive potential region exists in Fig. 3 -c).

DISCUSSION

In order to evaluate our models, distributions of polar and non-polar groups on the three dimensional structures were investigated using the program; POLDIAGNOSTICS developed by Baumann *et al.* [25] All the characteristic quantities are between the lowest and the highest limits, as shown in Table 2. It means that our models have similar characteristics of the polarity to that of native proteins.

Table 1. Suitable fragments for local loops of RT-RNase H domains.

Loops[1,2) and proteins	amino acid sequences of loops pre-loops	middle loops	post-loops	r.m.s.d.[3) (A)	
LM1	SLL	QEGQRK	AGA		
LH1	ANR	ETKLGK	AGY		
RNase H (13–22)	CLG	NPGP	GGY		
1GCR (6–17)	YED	RGFQGH	CYE	0.62	
LM2	VTT	ETEV	IWA		
RNase H (24–35)	AIL	RYRGRE	KTF		
1CMS (309–318)	VFD	RANN	LVG	0.27	
LM3	LKM	AEGK	KLN		
RNase H (56–66)	LEA	LKEHC	EVI		
9PAP (141–150)	FQL	YRGG	IFV	0.79	
LM4	KAL	FLPKR	LSI		RNase
H (109–118)	AAL	GQHQ	IKW		2RHV-
VP1(114–124)	KKL	ELFTY	VRF	0.55	
LM5	PGH	QKGHSAEA	RGN		RNase
H (122–130)	KGH	AGH	PEN		5API-
A (299–312)	LKS	VLGQLGIT	KVF	1.05	
LH2	AGY	VTNKGR	QKV		RNase
H (20–36)	GGY	GAILRYRGREK	TFS		3GRS
(367–378)	IPT	VVFSHP	PIG	0.56	
LH3	LQD	SGL	EVN		RNase
H (56–66)	LEA	LKEHC	EVI		1GOX
(117–125)	AST	GPG	IRF	0.75	
LH4	YAL	GIIQAQPDKS	ESE		RNase
H (73–102)	YVR	QGI ... PVK	NVD		1MBS
(33–48)	FKS	HPETLEKFDK	FKH	0.98	
LH5	EQL	IKKE	KVY		RNase
H (109–117)	AAL	GQH	QIK		2SNS
(64–73)	KMV	ENAK	KIE	0.55	

1) The local loops such as LM1 *etc.* are indicated in Fig. 1.
2) RNase H is *E.coli* RNase H [6]. Four-characters PDB codes [11] mean the following proteins. 1GCR; γ-II crystalline [15], 1CMS; Chymosin B [16], 9PAP; Papain [17], 2RHV-VP1; VP-1 subunit of Rhinovirus coat protein [18], 5API-A; A-chain of modified α_1-antitrypsin [19], 3GRS; Glutathione reductase [20], 1GOX; Glycolate oxidase [21], 1MBS; myoglobin [22], 2SNS; Staphylococcal nuclease [23]. Numbers in parentheses are residue numbers including pre- and post-loops.
3) r.m.s.d. is a root mean square deviation of backbone N, Cα, C′ atoms of the three pre-loop residues and the three post-loop residues between *E. coli* RNase H and the superposed local fragment structures in PDB.

Although the loop search method is now conventional to build a tertiary structural model of a homologous protein [26, 27], it is difficult to evaluate the correctness of the predicted loop structures. Especially for a long loop with a large deletion or insertion of amino acids, it is impossible to select a specific loop among lots of possible fragments without any experimental information. In the present model building, the loops of LM5 and LH4 shown in Fig. 1 are too long to construct unambiguous models. The r.m.s.d. values of LM5 and LH4 in Table 1, which were calculated for pre- and post- six residues between *E. coli* RNase H and the superposed fragments, are also high. That is, the reliability of these loop structures is not tolerable for further discussion. However, as far as one assumes that the backbone structures of the framework do not deviate much from the original structure of *E. coli* RNase H, conformations of the framework are thought to be definite enough for discussion of the structure-function relationships of the RT-RNase H domains.

In the previous structural [6,7] and protein engineering [8] studies of *E. coli* RNase H, structurally and functionally essential residues have been revealed. Most of them are found to be well conserved in the amino acid sequences in RT-RNase H domains. In the current models, those residues are structurally conserved, having the same importances as in *E. coli* RNase H. Three acidic residues Asp 10, Glu 48 and Asp 70 in *E. coli* RNase H construct the active site, and His 124 and Asp 134 are in its neighborhood. A hydrogen bond between the side chain of Asn 130 and the backbone carbonyl Ser 68 supports a close contact between βD and αV. Arg 46 and Asp 102 make a salt bridge between αI and αIV. All these residues are retained in the similar tertiary positions in MuLV and HIV-1 RT-RNase H models as shown in Figs. 2 -a) and -b). Hydrophobic contact between two parallel helices, αI and αIV, are also observed in these models. Other amino acid residues forming the hydrophobic core are sometimes replaced with Ser or Thr, but they are mostly left as hydrophobic or aromatic residues. These conserved residues are marked in Fig. 1.

It has been pointed out [6, 7, 9] that the charge distribution of *E. coli* RNase H has an essential role in recognition of the

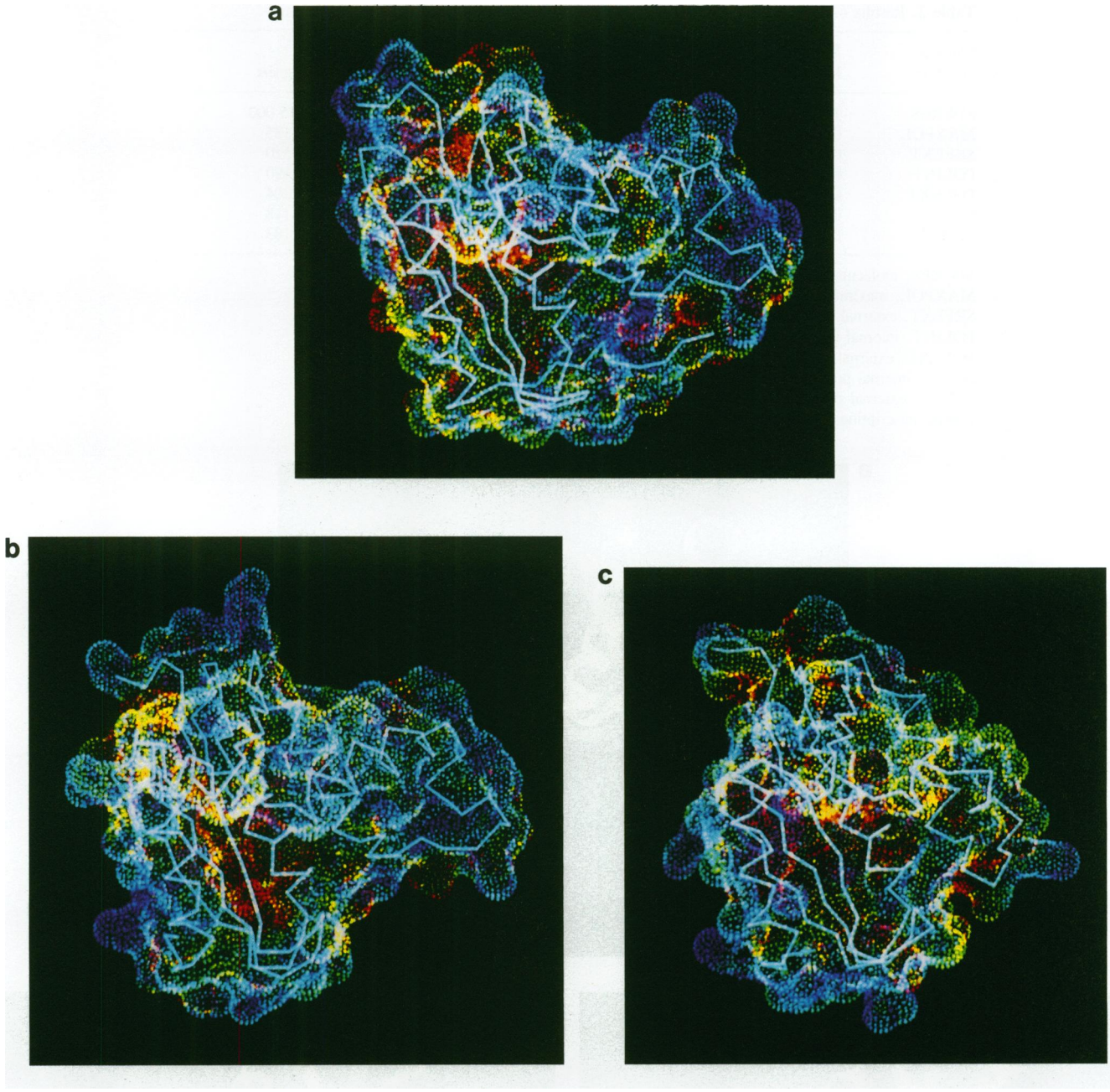

Figure 3. Electrostatic molecular surfaces of (a) *E. coli* RNase H without the bound Mg^{2+} ion, (b) the model of MuLV RT-RNase H domain and (c) the model of HIV-1 RT-RNase H domain in the same view as Fig. 2. Colors show electrostatic potential values: red, below −0.1 V and blue above +0.1 V. Other color codes correspond to values between +0.1 V and −0.1 V sequentially following the spectrum from blue to red. White lines are backbone Cα structures.

substrate hybrid. Especially, positive charges in the minor domain shown in Fig. 3 -a) may interact with phosphate groups of DNA-RNA hybrid. In the current model of MuLV RT-RNase H, not only the backbone structure but also the electrostatic features are well retained as shown in Fig. 3 -b). In fact, deletion mutants of MuLV RT including only RNase H domain were found to have RNase H activity [28]. On the contrary, as shown in Figs. 1, 2 -b) and 3 -c), the model of HIV-1 RT-RNase H has a large gap in this area, which make no significantly positive potential surface there. From deletion studies of RT from HIV-1, RNase H activity is considered not to be associated with the C-terminal domain alone, but with many parts in the whole RT [29]. Our model of HIV-1 RT-RNase H suggests that the DNA-polymerase domain may help the hybrid recognition by giving electrostatically positive potential in the neighborhood of the minor domain.

As shown in Fig. 4 -a), a scheme of the substrate binding to *E. coli* RNase H has been proposed based upon our previous ^{15}N-oriented NMR experiments, where backbone protons of the enzyme affected by titration of the substrate hybrid were analyzed [9]. The side chains of the active site residues are also shown by red. Here, lower right region of the enzyme interacting the 21-mer DNA-RNA hybrid is the minor domain. This complex model is essentially similar to that built by Yang *et al.* [7] When *E. coli* RNase H structure in that complex is superposed to the models of RT-RNase H domains, the hybrid can interact with both MuLV and HIV-1 RT-RNase H domains in the similar

Table 2. Results of POLDIAGNOSTICS [25] for two model structures of RT-RNase H domains.

Measure quantities	MuLV RT-RNase H	HIV-1 RT-RNase H	in 150 proteins lowest	mean	highest
MWRES	109.413	109.473	97.000	109.000	115.000
MAXPOL	0.176	0.174	0.163	0.173	0.185
SRFEXT	0.447	0.446	0.320	0.400	0.520
POLINT	0.186	0.185	0.158	0.172	0.190
POLEXT	0.160	0.155	0.147	0.172	0.204
SCINT	0.110	0.110	0.074	0.098	0.118
SCOUT	0.139	0.130	0.115	0.143	0.183

MWRES, molecular weight per residue (unit is dalton/res);
MAXPOL, maximum polar fraction (e);
SRFEXT, external surface area per molecular weight (A^2/dalton);
POLINT, internal polar fraction (e);
POLEXT, external polar fraction (e);
SCINT, internal polar fraction for side chains (e);
SCEXT, external polar fraction for side chains (e).
Precise descriptions of these quantities are in ref. [25].

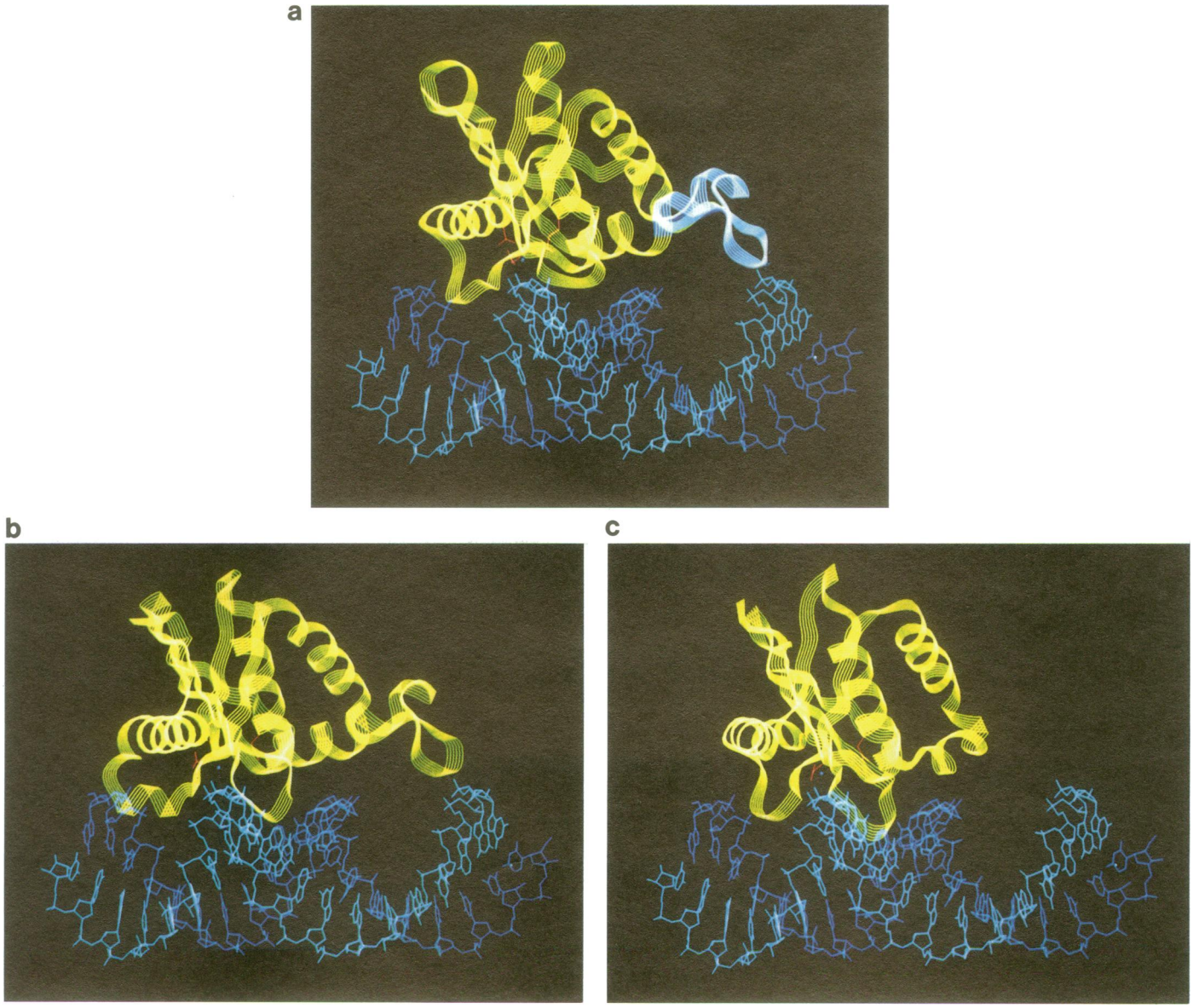

Figure 4. A 21-mer DNA-RNA hybrid complex model with (a) *E. coli* RNase H, (b) the model of MuLV RT-RNase H and (c) the model of HIV-1 RT-RNase H. Yellow ribbons show the backbone structures of the enzymes. The minor domain in *E.coli* RNase H is shown by a white ribbon in fig. 4. -a). Cyan wires are the RNA strands, whose 5'-ends are at the right sides of the figures. Blue wires are the DNA strands. Three carboxylates of Asp 10, Glu 48 and Asp 70 side chains in *E.coli* RNase H and associated residues in RT-RNase Hs, which form the RNase H active sites, are also shown by red wires. Here, the enzymes are looked down from the tops of Figs. 2-a) and -b).

manner to *E. coli* RNase H. Those complex models of RT-RNase H domains are illustrated in Figs. 4 -b) and -c), respectively. Except the suppressed interaction between the substrate hybrid and the minor domain of HIV-1 RT-RNase H, the interaction scheme seems essentially the same for all these RNase Hs.

Oyama *et al.* [3] show that RT RNase Hs cleave the RNA strand of the DNA-RNA hybrid, leaving an hybrid with 5−10 base pairs, with and without DNA polymerization. For MuLV RT-RNase H, the associated base pair number is 5. In case of *E. coli* RNase H, the number is only two. Recent study by Schatz *et al.* [4] reveals that the base pair number for HIV-1 RT-RNase H is 7. These base pair numbers depend upon the species of RTs, and they may reflect varieties of the interaction schemes between the domains of DNA-polymerase and RNase H [3]. Oyama *et al.* [3] and Schatz *et al.* [4] proposed a tight coupling model between the two different activities of RTs, where cDNA synthesis and RNase H hydrolysis of the template RNA can proceed simultaneously. In their structural model, the DNA polymerase active site in RTs is located at the 5′-side of RNA strand. When their tight coupling model is assumed, the active site of DNA-polymerase domain should be close to the minor domain of RNase H, about 10 base pairs upstream from the RNase H active site, as seen in Figs. 4 -a) and -b). Considering that RNase H domain alone of cloned HIV-1 RT does not retain RNase H activity [29], the interaction between polymerase and RNase H domains around the minor domain is thought to be crucial for RT activities of HIV-1.

In conclusion, we built tertiary models of MuLV and HIV-1 RT-RNase H domains, which have many similar structural features to those of *E. coli* RNase H. DNA polymerase domain of RT is considered to interact mainly with the minor domain of RNase H, giving rise a concomitant RNA cleavage with DNA polymerization. In case of HIV-1 RT-RNase H, the DNA-polymerase domain may help substrate recognition of RNase H domain.

ACKNOWLEDGEMENTS

The authors wish to thank C. Sander for their programs DSSP and POLDIAGNOSTICS. The authors wish to thank S. Kanaya and T. Nakai for discussions.

REFERENCES

1. Baltimore, D. (1985) *Cell*, **40**, 481−482.
2. Crouch, R. J. (1990) *The New Biologist*, **2**, 771−777.
3. Oyama, F., Kikuchi, R., Crouch, R. J. and Uchida, T. (1989) *J. Biol. Chem.*, **264**, 18808−18817.
4. Schatz, O., Mous, J. and Le Grice, S. F. J. (1990) *EMBO J.*, **9**, 1171−1176.
5. Doolittle, R. F., Feng, D. -F., Johnson, M. S. and McClure, M. A. (1989) *Quart. Rev. Biol.*, **64**, 1−30.
6. Katayanagi, K., Miyagawa, M., Matsushima, M., Ishikawa, M., Kanaya, S., Ikehara, M., Matsuzaki, T. and Morikawa, K. (1990) *Nature (London)*, **347**, 306−309.
7. Yang, W., Hendrickson, W. A., Crouch, R. J. and Satow, Y. (1990) *Science*, **249**, 1398−1405.
8. Kanaya, S., Kohara, A., Miura, Y., Sekiguchi, A., Iwai, S., Inoue, H., Ohtsuka, E. and Ikehara, M. (1990) *J. Biol. Chem.*, **265**, 4615−4621.
9. Nakamura, H., Oda, Y., Iwai, Ohtsuka, E., S., Kimura, S., Katsuda, C., Katayanagi, K., Morikawa, K., Miyashiro, H. and Ikehara, M., submitted.
10. Jones, T. A. and Thirup, S. (1986) *EMBO J.*, **5**, 819−822.
11. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. and Tasumi, M. (1977) *J. Mol. Biol.*, **112**, 535−542.
12. Weiner, S. J., Kollman, P. A., Nguyen, D. T. and Case, D. A. (1986) *J. Comput. Chem.*, **7**, 230−252.
13. Nakamura, H. and Nishida, S. (1987) *J. Phys. Soc. Jap.*, **56**, 1609−1622.
14. Connolly, M. L. (1983) *J. Appl. Crystallogr.*, **16**, 548−558.
15. Summers, L., Wistow, G., Narebor, M., Moss, D., Lindley, P., Slingsby, C., Blundell, T., Bartunik, H. and Bartels, K. (1984) *Pept. Protein Rev.*, **3**, 147−168.
16. Gilliland, G. L., Winborne, E. L., Nachman, J. and Wlodawer, A. (1990) *PROTEINS*, **8**, 82−111.
17. Kamphuis, I. G., Kalk, K. H., Swarte, M. B. A. and Drenth, J. (1984) *J. Mol. Biol.*, **179**, 233−256.
18. Arnold, E., Vriend, G., Luo, M., Griffith, J. P., Kamer, G., Erickson, J. W., Johnson, J. E. and Rossmann, M. G. (1987) *Acta Crystallogr.*, A43, 346−361.
19. Loebermann, H., Tokuoka, R., Deisenhofer, J. and Huber, R. (1984) *J. Mol. Biol.*, **177**, 531−556.
20. Karplus, P. A. and Schulz, G. E. (1987) *J. Mol. Biol.*, **195**, 701−729.
21. Lindqvist, Y. (1989) *J. Mol. Biol.*, **209**, 151−166.
22. Scouloudi, H. and Baker, E. N. (1978) *J. Mol. Biol.*, **126**, 637−660.
23. Cotton, F. A., Hazen Jr., E. E. and Legg, M. J. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 2551−2555.
24. Kabsch, W. and Sander, C. (1983) *Biopolymers*, **22**, 2577−2637.
25. Baumann, G., Frommel, C. and Sander, C. (1989) *Protein Engineering*, **2**, 329−334.
26. Sutcliffe, M. J., Haneef, I., Carney, D. and Blundell T. L. (1987) *Protein Enginerring*, **1**, 377−384.
27. Martin, A. C. R., Cheetham, J. C. and Rees, A. R. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 9268−9272.
28. Tanese, N. and Goff, S. P. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 1777−1781.
29. Prasad, V. R. and Goff, S. P. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 3104−3108.