Volume 324, number 1, 15–21 FEBS 12525 June 1993

Three-dimensional structural resemblance between the ribonuclease H and connection domains of HIV reverse transcriptase and the ATPase fold revealed using graph theoretical techniques

Peter J. Artymiuk[a], Helen M. Grindley[b], Kiran Kumar[a], David W. Rice[a] and Peter Willett[b]

[a]*Krebs Institute, Departments of Molecular Biology and Biotechnology and* [b]*Information Studies, University of Sheffield, Sheffield, S10 2TN, UK*

Received 5 April 1993

Using 3D searching techniques based on algorithms derived from graph theory, we have established two previously unreported structural similarities involving the ribonuclease H (RNase H) domain of HIV-1 reverse transcriptase (RT). First, we report that there is a strong similarity between the 3D folds of the RNase H domain of RT and the 'ATPase folds' of hexokinase, the 70 kDa heat-shock cognate protein and actin. Like RNase H, these enzymes are involved in nucleotide binding and metal ion-catalysed cleavage of a phosphodiester bond. Similarities of the folding motif and the position of the metal-binding site in these enzymes suggest possible functional analogies and evolutionary relationships with RNase H. Second, we find there is a strong resemblance between the folds of the RNase H domain and of the p66 and p51 'connection' domains of RT. It is possible that this striking similarity within the RT structure indicates a possible ancestral gene doubling event. The similarity may also indicate that the connection domains possess functional roles in addition to those previously suggested, and they may therefore represent a further target for the design of therapeutic agents.

Structural similarity; Graph theory; HIV; AIDS; Reverse transcriptase; Ribonuclease H; ATPase fold

1. INTRODUCTION

Viral reverse transcriptases (RTs) are multifunctional enzymes which catalyse the transcription of a single-stranded retroviral RNA template into a single strand of DNA, and single-stranded DNA into double-stranded DNA which is then subsequently incorporated into the host cell's genome [1]. This process is unique to retroviruses, and the RT of the HIV virus is therefore a prime target for potential anti-AIDS drugs such as AZT (3-azido-deoxythymidine) and ddI (dideoxyionisine), which both operate by incorrectly terminating the DNA strand as it is synthesized by RT [2]. Major medical problems have arisen, however, because mutations in RT occur rapidly and confer resistance to these drugs (see [3]).

HIV RT is initially produced from the viral polyprotein gene product as a 66 kDa polypeptide with both polymerase and ribonuclease H (RNase H) activity. Two of these p66 polypeptides associate to form a homodimer, but subsequently the RNase domain is cleaved from the C-terminus of one of the p66 chains to produce a 51 kDa polypeptide (p51) [4,5]. The final functional RT molecule is thus a heterodimer consisting of one p66 and one p51 chain. The final protein appears to have only one functional polymerase site, and one RNase H site which is used for a variety of distinct RNA cleavage events [6], including the destruction of the RNA template after transcription and the removal of RNA primer.

The structure of the HIV RNase H domain has been determined by Davies et al. [7], and a low resolution structure of a ternary complex of RT with DNA and an antibody Fab fragment has been solved by Arnold et al. [8]. However, the understanding of RT structure and function took a major step forward recently with the determination of the crystal structure of a complete HIV-1 RT at 3.5 Å resolution by Kohlstaedt et al. [9], which has revealed the chain fold of the entire protein, the mode of association of p66 and p51, the location of the polymerase active site, and suggested the site of interaction of nucleic acid with RT. The p66 chain can be divided into five domains [9]: the first three, called the 'fingers', 'palm' and 'thumb' domains, grip the proposed position of the DNA–RNA duplex in a manner analogous to the equivalently named domains in the Klenow fragment of DNA polymerase I [10]; the fourth domain is called the 'connection' domain and leads into the RNase H domain. The p51 chain contains the fingers, palm, thumb and connection domains, but with a

Correspondence address: P.J. Artymiuk, Krebs Institute, Department of Molecular Biology and Biotechnology, University of Sheffield, Sheffield, S10 2TN, UK. Fax: (44) (742) 728 697.

Abbreviations: RT, reverse transcriptase; RNase, ribonuclease, HIV, human immunodefiency virus; AIDS, acquired immune deficiency syndrome; PDB, Protein Data Bank (Brookhaven); 3D, three-dimensional; SSE, secondary structure element.

drastic quaternary structural rearrangement of the latter two domains relative to the first two, which means that there is no active site cleft in p51.

There is considerable interest in the relationship of the RT structure to other known structures, as structural analogies may provide information for the rational design of drugs, and indeed Kohlstaedt et al. [9] have pointed out the important similarity between the palm domains of RT and of the Klenow fragment [10]. In this paper we use methods derived from graph theory [11–13] to identify two intriguing and previously unrecognized structural similarities: (i) between the 3D folds of the RNase H domain of RT and the 'ATPase folds' of hexokinase, the 70 kDa heat-shock cognate protein (HSC70) and actin, and (ii) a structural resemblance within the RT structure itself, between the connection domain and the RNase H domain.

2. MATERIALS AND METHODS

2.1. *Linear representations of helices and strands*

Regions of helix and strand in proteins in the Protein Data Bank (PDB) [14,15] (April 1992 release) were assigned using the algorithm of Kabsch and Sander [16]. The position and direction of each secondary structure element (SSE) was then approximated by a vector in 3D space which corresponds to the axis of an idealized helix or strand superposed on the real helix or strand by least squares. The torsional angles, closest approach distances and distances between midpoints of each pair of SSEs within each protein in the PDB are stored in a database as a labelled graph. The nodes of the graph are the linear representations of the SSEs, and the edges of the graph the distances and angles between them [11].

2.2. *Detection of subgraph isomorphism*

The PROTEP program [12] (Tripos Associates) uses a maximal common subgraph algorithm [17] to match the query nodes, i.e SSEs in the present context, to the structure nodes by looking at the relationships (graph edge values, within specified tolerances) between them. This permits the rapid location of any structural overlaps between the query structure and any of the other proteins in the PDB, and output is interfaced to the FRODO graphics program [18]. β-Sheet topology searches were performed using the subgraph isomorphism program POSSUM [11]. The search programs have been extensively tested against a variety of known motifs, including the trypsin, azurin and globin families, and shown to operate correctly and effectively [12,13].

2.3. *Note on coordinate sets used*

2.3.1. HIV RT

Coordinates of HIV RT have been deposited in the PDB by Kohlstaedt et al. [9], where they are presently available as prerelease entry 1HVT in the October 1992 release. The entry consists of C-α's only, many loops have been removed pending more detailed interpretation, and in the preamble to the coordinate set Kohlstaedt et al. explain that the chain has been arbitrarily renumbered starting from '1' at the N-terminus of p66 and from '601' at the N-terminus of p51 with missing loops represented by a gap of only 1 in the numbering. For ease of comparison of coordinates in this paper, we therefore use this arbitary renumbering system for RT, but precede the residue number by an asterisk to indicate this (e.g. *370). The result of this renumbering is that the finger and palm domains, which correctly consist of residues 1–240 in the sequence, are renumbered approximately *1–*198 (p66) and *601–*775 (p51); the thumb domains, correctly 241–315, are renumbered *199–*279 (p66) and *776–*824 (p51); the connection domains, correctly 315–426, are renumbered *280–*368 (p66) and *833–*918 (p51); and the RNase H domain, correctly 427–563, is numbered *370–*502 (present in p66 only).

Nevertheless, it is important to observe that the authors emphasize in the preamble to the PDB release that they are confident that the direction of the chain trace is correct everywhere. Consequently the conclusions of this paper which are concerned with the organization of α-helices and β-strands within the fold, and not in the precise locations of side chains, are not in doubt.

2.3.2. HIV ribonuclease H

The coordinates for the HIV ribonuclease H are deposited in the PDB by Davies et al. [7] as deposition 1HRH. The coordinates consist of main chain and side chain atoms, but with certain loops omitted. The residues are correctly numbered according to the numbering in the complete RT p66 chain and commence with residue Tyr-427 and end with Ala-554. These coordinates were used by Kohlstaedt et al. [9] to interpret the RNase H region of their RT structure. There are two symmetry-related molecules in the crystallographic asymmetric unit, and we have chosen chain 'B' for these comparisons. The homologous *E. coli* RNase H [19] has a very similar 3D structure to the HIV enzyme [7] and is deposited as 1RNH.

3. RESULTS

Two search patterns were constructed, consisting, respectively, of the secondary structure elements of the RT p66 (15 helices, 23 strands) and RT p51 (10 helices, 16 strands) chains of HIV-1 RT (PDB code 1HVT, from the October 1992 prerelease of the PDB). The computer program PROTEP [12,13] was used to search the April 1992 release of the Protein Data Bank for similar 3D structures. CPU times for the search were 24 min (p66) and 14 min (p51) on a Silicon Graphics Indigo R4000 workstation. We monitored all structural similarities consisting of at least six SSEs arranged in 3D space within tolerances of 30° on inter-SSE angles and of the lesser of 40% or 4 Å on inter-SSE closest approach distances. As a further constraint, only those structural matches where SSEs occurred in the same sequence order in both the search protein and the PDB proteins were monitored.

Because p66 contains the RNase H domain, we were not surprised to find structural overlaps ('hits') between p66 and the separately solved structure of HIV ribonuclease H [7] (PDB code 1HRH), or with the homologous RNase H from *E. coli* [19] (1RNH), the latter having 25% sequence identity with the HIV enzyme. Surprisingly, however, we found that the connection domains of both p66 and p51 also strongly resembled the RNase H structures, indicating a previously unreported structural repeat within the RT structure itself. In addition, although Yang et al. [19] reported no substantial similarity between RNase H and any other known structures, significant overlaps were found between the RNase H (and connection) domains with the Ia and IIa domains of hexokinase (PDB codes: 1HKG and 2YHX) [20] and of the ATPase fragment of HSC70 (PDB code: 1HSC) [21]. The Ia and IIa domains of the latter two proteins are known to have a common ATPase fold which they also share with actin [22] and glycerol kinase [23] (neither present in the April 1992 release of the PDB, and therefore not found in our search). These

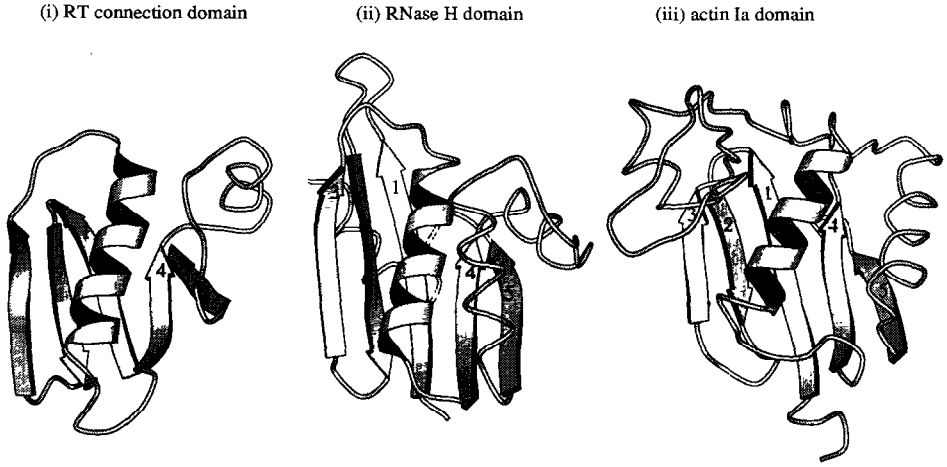

Fig. 1. Chain traces (produced using the program Molscript [24]) of (i) the connection domain of RT, (ii) the RNase H domain of RT, and (iii) the Ia domain of actin. The helices and strands equivalenced in our study are represented as coiled ribbons and sequentially numbered arrows, respectively, the non-equivalent parts of the structure are shown as smoothed C-α trace.

similarities are tabulated in Table I, and are illustrated in Figs. 1–3.

4. DISCUSSION

4.1. *Comparison of the RNase H domain with the ATPase fold*

It is clear that the five-stranded β-sheet in RNase H and the five-stranded sheet in the ATPase fold are topologically identical (i.e. they share the same directionality and connectivity of β-strands). Furthermore, this is an uncommon β-sheet motif: a topological search of the PDB using the POSSUM program [11] reveals that no other proteins contain this strand arrangement. The overlap of the ATPase fold proteins with RNase H also includes the α-helix which runs between strands 3 and 4 (see Fig. 1 and Table I). The agreement in 3D space is good: thus, for example, 45 C-α atoms from the core regions of the five strands and the helix in RNase H (1HRH) and in the hexokinase (2YHX) Ia domain superpose with an RMS error of 1.76 Å, however the helices in both structures between strands 4 and 5 do not superpose well, and the ATPase domains contain a link to an extra domain (Ib or IIb) after strand 3.

Table I

Residue numbers of the matched helices and strands in the connection domains of RT p66 and RT p51 [9], RT RNase H [7] and the hexokinase Ia domain [20]

PDB entry	RT p66 connect 1HVT	RT p51 connect 1HVT	RT RNase H 1HRH	Hexokinase domain Ia 2YHX
Strand 1	*281–287	*835–840	442–447	63–69
Strand 2	*290–294	*843–847	452–460	74–79
Strand 3	*303–307	*855–861	463–470	90–93
Helix A	*313–327	*871–886	473–489	109–123
Strand 4	*336–340	*891–895	491–498	130–136
Strand 5	*361–366	*913–917	529–536	185–189

Comparable equivalences are found in the hexokinase IIa domain, and the actin and HSC70 Ia and IIa domains (data not shown). The asterisks indicate the use of the preliminary numbering system adopted by Kohlstaedt et al [9] in their PDB deposition 1HVT, in which the residues of the connection domain (correctly 315–426) are renumbered *280–*368 (p66) and *833–*918 (p51) (see section 2.3.1.).

Although there is no general sequence similarity between hexokinase, the HSC70 protein and actin, the strong structural similarities between their Ia and IIa domains has led to speculation about their probable evolutionary relatedness [21–23]. Moreover, the internal similarity between the Ia and IIa domains themselves has been taken as indicative of a possible gene doubling event having given rise to these structures [22], implying a monomeric ancestral molecule. Indeed, Naharro et al. [26] have shown that the v-*fgr onc* gene product of the Gardner–Rasheed feline sarcoma virus consists of a single ATPase domain of actin fused to a tyrosine-specific protein kinase.

It appears that RNase H also represents a monomeric version of this fold, which in the case of the *E. coli* RNase H [7] can function independently of other proteins. Moreover, not only is there a similarity in the folding motif, but in both the ATPases and in RNase H there is a functionally important aspartate residue which ligates divalent metal ions (e.g. Asp-11 in the actin Ia domain and Asp-154 in the actin IIa domain bind a Ca^{2+} ion, and Asp-443 in RNase H ligates one of a pair of divalent metal ions), and which is located at the same position in strand 1 of the sheet. It is also interesting to note that this metal site is located in a position in the β-sheet where strands 1 and 4 diverge from each other to form a cleft in the β-sheet (see Fig. 1) in both the RNase H and the ATPase fold structures. It can be argued that the RNase H has a related function to ATPase fold molecules in that they are all involved

Table II

(a) Rotation matrix used to rotate 1HRH (B chain) onto the p66 connection domain of 1HVT

$$\underline{X}' = \begin{matrix} -0\,33639 & -0.87910 & -0.33768 \\ 0.40461 & -0.45871 & 0.79112 \\ -0.85037 & 0.12950 & 0.51000 \end{matrix} \cdot \underline{X}_{HRH} + \begin{matrix} 162\,42 \\ -29.52 \\ 56.99 \end{matrix}$$

(b) Rotation matrix used to rotate 2YHX (Ia domain) onto the p66 connection domain of 1HVT

$$\underline{X}' = \begin{matrix} -0.05777 & 0.00938 & -0.99829 \\ 0.10294 & -0.99457 & -0.01530 \\ -0.99301 & -0.10365 & 0.05649 \end{matrix} \underline{X}_{YHX} + \begin{matrix} 111.55 \\ -52.97 \\ 143.80 \end{matrix}$$

$\underline{X}_{HRH}$ and $\underline{X}_{YHX}$ are column vectors representing the 1HRH and 2YHX coordinates, respectively, and $\underline{X}'$ the same coordinates superposed on the RT p66 connection domain structure. Equivalent matrices can be calculated for the superposition onto the RT p51 connection domain (data not shown).

in metal-dependent cleavage of a phosphodiester bond. Thus it is not unreasonable to speculate on the existence of an ancient evolutionary connection between these classes of molecules. Nevertheless, a note of caution must be sounded: it is quite possible that the similarities we have established between RNase H and the ATPase fold are due to the processes of convergent evolution towards a common structure that catalyses out analogous reactions. Such convergence is known to occur at the level of groups of side chains, as for example in the catalytic triad in the trypsin and subtilisin families of serine proteases and lipases [27].

4.2. *Similarity between connection domain and the RNase H and ATPase folds*

The situation becomes more intriguing however, when one considers the other similarity we have detected, namely that the motif that is common to RNase H and the ATPase fold proteins is also present in the RT connection domain (see Figs. 1 and 2). The equivalenced motif of five strands and a single helix in the RNase H and connection domains have identical topology and superpose well in 3D with an RMS error of 1.77 Å over 48 core C-α-atoms (Fig. 3a–d). Sequence comparisons reveal no significant similarity between sequences of the RNase H and connection domains of RT, and thus there can be no direct evidence of any ancestral gene duplication event in the evolution of the RT protein. Nevertheless, particularly as the shared motif is an unusual one (see above), it seems probable that the similarity is not coincidental.

A number of 3D structural duplications have been observed within protein structures, and often there is no evidence of sequence similarity between the structurally similar regions. A well-known example is the ATPase-fold discussed above, and numerous other examples include rhodanase, the Rossmann fold and the serine proteases [28]. In spite of the lack of internal sequence similarity, it is argued that such internal similarities have arisen from ancient gene doubling events. The rationale given for this is that 3D structure is conserved more strongly than amino acid sequence over long evolutionary periods, and so the structural similarity remains long after significant sequence similarity has been lost [29].

If the similarity between the RT connection and RNase H domains is due to a gene doubling event, then

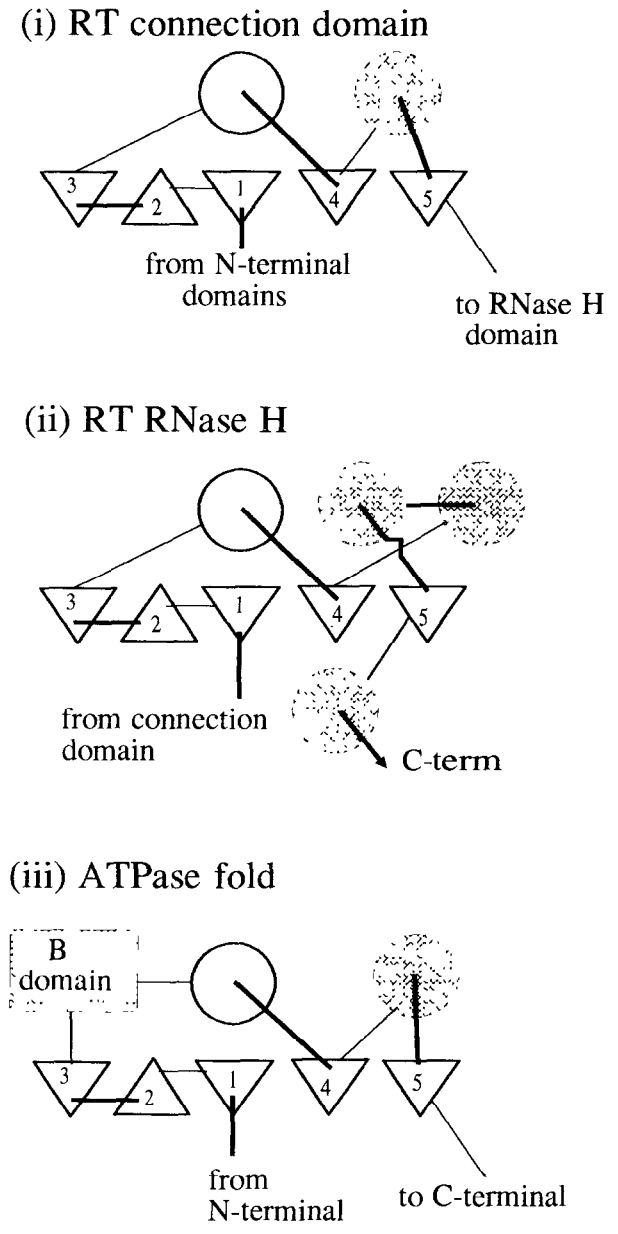

Fig. 2 Topological diagrams [25] of (i) the connection domain of RT, (ii) the RNase H domain of RT, and (iii) Ia domain of hexokinase. Circles represent α-helices and triangles represent β-strands (apex down indicates the strand is viewed from the C-terminus). Open circles and triangles indicate those secondary structure elements that were found by the PROTEP search to superpose in 3D, and shaded ones those that do not. The strands in the 5-stranded β-sheets are numbered according to their order in the sequence.

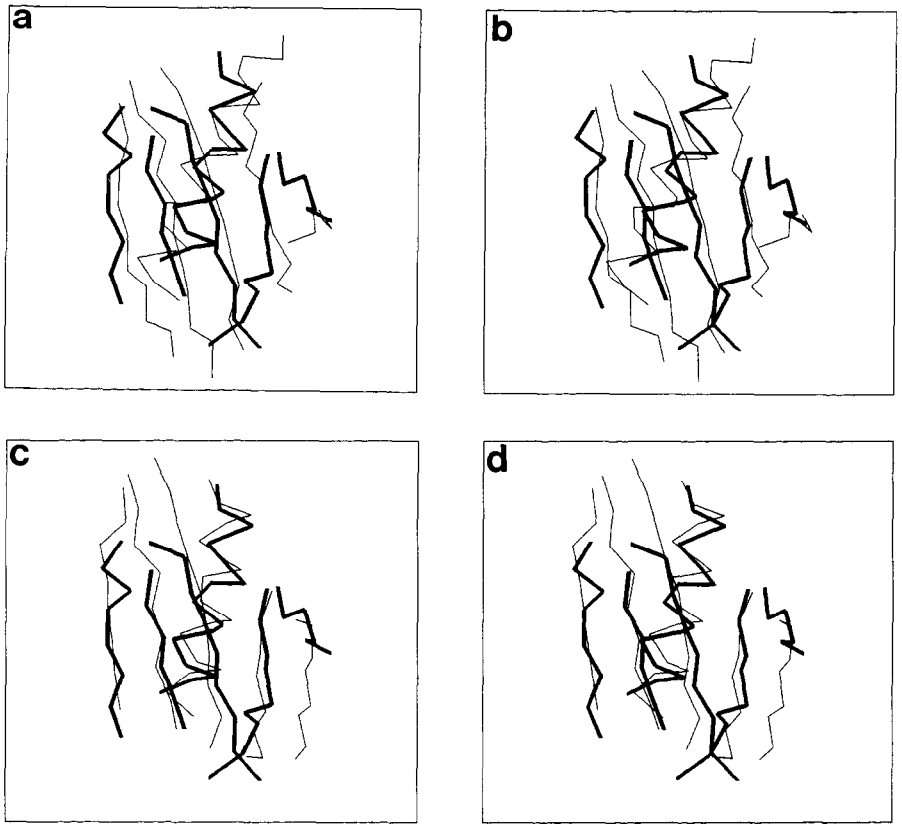

Fig. 3. Stereodiagrams showing the superposition of C-α's of the hexokinase Ia domain (a and b, fine lines) and the RT RNase H domain (c and d, fine lines) on the RT p66 connection domain (bold lines). Only equivalenced helices and strands are shown for clarity: the non-equivalent parts of the structures are indicated in Figs. 1 and 2. The IIa domain of hexokinase and the Ia and IIa domains of the HSC70 ATPase fragment and actin are similar to the hexokinase Ia domain and is not shown. The p51 connection domain is similar to the p66 connection domain and is not shown.

there are a number of possible scenarios for the sequence of evolutionary events, which must, however, take account of the fact that the bacterial RNase H's are known to have sequence homology with the retroviral RNase H domain [7,19]. There are three main possibilities: (i) a bacterial RNase H was the ancestor of the retroviral enzyme, which underwent a gene doubling event to produce a connection domain, and at some point became associated with the retroviral polymerase; (ii) a gene doubling event in retroviruses gave rise to the connection domain–RNase H domain pair; or (iii) the connection domain–RNase H pair evolved from an ancestral ATPase molecule that already had the gene-duplicated ATPase fold. In proposal (ii), and perhaps in proposal (iii), the bacterial RNase H's would be descended from the retroviral enzyme, which would be not be inconsistent with the fact that, whereas RNase H is essential for retroviral replication, it is a non-essential enzyme in *E. coli* [30]. Proposal (iii) could be consistent with either a bacterial or a viral origin for the RNase H enzymes. An objection to proposal (iii) is that there is no similarity between the relative dispositions of the RNase H and connection domains in RT, and of the Ia and IIa domains in hexokinase, HSC70 and actin. In fact, this is not necessarily an overwhelming objection: it is clear from comparisons of the relative dispositions of equivalent domains in the p66 and p51 chains of RT that massive tertiary structural changes must take place in RT when the heterodimer assembles [9]. In this context it would be most interesting to observe the structure of the p66 homodimer before cleavage of the second RNase H domain.

Clearly the evolutionary sequence of events must, for the moment at least, remain a matter for conjecture and controversy. Nevertheless, given the structural similarity with RNase H, it is of considerable interest to consider the possible roles of the two connection domains. In addition to their structural role in forming monomer–monomer contacts in RT, the connection domains associate together to form the bottom of the DNA–RNA duplex binding site and are positioned to interact with the duplex backbone approximately midway (i.e. one turn of DNA) between the DNA polymerase site and the RNase H site [9]. In view of the structural similarity between the metal-binding RNase H and ATPase proteins, it will be of interest to see if the contin-

uing crystallographic work [9] reveals any metal binding sites in these domains which could be involved in stabilizing the negatively charged sugar–phosphate backbone of the nucleic acid. Tandem association of similar protein subunits with DNA is seen in the structures of the Zif/ 268 zinc finger protein [31], but it is doubtful whether the three structurally similar domains in RT (p51 connection domain, p66 connection domain, and p66 RNase H domain) can be regarded as interacting with the nucleic acid chain in an analogous manner. This is because (in the native structure) the orientation of each of the three domains is different with respect to the expected nucleic acid binding site; however, there have been preliminary reports [9] of conformational changes in RT upon binding nucleic acid, which may alter this perception.

In view of the similarity in structure of the connection domains to RNase H, it is also worthwile to consider the RNase activities associated with RT. In addition to DNA polymerase and RNase H activity, the presence of an additional RNase D activity in RT, associated with cleavage of RNA primer, has been reported [32], although this finding has been questioned by others [33]. The presence of the two connection domains with structures analogous to that of RNase H may provide candidate sites for such activity if it is confirmed. In addition, the RNase H activity of RT is important at several stages of viral genome replication, including removal of the RNA template after first strand DNA synthesis, cleavage of the host tRNA primer, generation of a specific oligopurine ribonucleotide primer from which synthesis of second strand DNA is initiated, and its subsequent removal [6]. Peliska and Benkovic [34] identify two RNase H activities in RT which are required for removal of RNA fragment annealed to the nascent DNA strand: (i) the template RNA is degraded at a locus 18-19 bases away from the polymerase site, and (ii) a kinetically distinct RNase H activity is associated with DNA strand transfer which releases RNA fragments a maximum of 14 bases long. A difficulty with this, however, is that the RNase H site is not 14 base pairs, but 18-19 base pairs from the polymerase site, and Peliska and Benkovic [34] are obliged to suggest that the duplex must alter its mode of binding in order to achieve the observed activity. In this connection, it is intriguing to observe that if the p51 connection domain had RNase H activity, it would be positioned in approximately the correct position some three-quarters of the way between the DNA polymerase and RNase H active sites. Moreover, unlike the p66 connection domain which is facing the wrong way from the DNA–RNA duplex, the orientation of the p51 connection domain relative to the DNA–RNA binding site would not be inconsistent with that of the RNase H domain with respect to the duplex. We therefore suggest that it may be valuable to study the connection domains of the RT using site-directed mutagenesis with a view to confirming or refuting the presence of possible RNase activity, which, if present, would provide another potential target for drug design.

In conclusion, given our present lack of knowledge concerning this important therapeutic target, our analysis reveals a number of new and potentially important aspects of the structure of RT. These relate both to possible evolutionary origins of some of the functions of RT, and to potential candidates for the locations of additional functional sites involved either in catalysis or in stabilization of nucleic acid.

Acknowledgements. This work was supported by Science and Engineering Research Council, The Wellcome Trust and Tripos Associates Inc. The Krebs Institute is a designated centre of the SERC Molecular Recognition Initiative. We acknowledge the use of the SERC SEQNET facility. P.J.A. is a Royal Society University Research Fellow, and D.W.R. is a Lister Institute Research Fellow.

REFERENCES

[1] Goff, S.P. (1990) J. AIDS 3, 817–831.
[2] Mitsuya, H., Yarchoan, R. and Broder, S. (1990) Science 249, 1533–1544.
[3] Wlodawer, A. (1992) Science 256, 1766.
[4] Lightfoote, M.M., Colligan, J.E., Folks, T.M., Fauci, A.S., Martin, M.A. and Vankatesean, S. (1986) J. Virol. 60, 771–775.
[5] Veronese, F.M., Copeland, T.D., DeVico, A.L., Rahman, R., Orcesz, I S., Gallo, R.C. and Sarngadharan, M.G. (1986) Science 231, 1289–1291.
[6] Gopalakrishnan, V., Peliska, J.A and Benkovic, S.J. (1992) Proc. Natl. Acad. Sci. USA 89, 10763–10767.
[7] Davies, J.F., Hostomska, Z., Hostomsky, Z., Jordan, S.R. and Matthews, D.A. (1991). Science 252, 88–95.
[8] Arnold, E., Jacobo-Molina, A., Nanni, R.G., Williams, R.L., Lu, X., Ding, J., Clark Jr., A.D., Zhang, A., Ferris, A.L., Clark, P., Hizi, M. and Hughes, S.H. (1992) Nature 357, 85–89.
[9] Kohlstaedt, L.A., Wang, J., Friedman, J.M., Rice, P A. and Steitz, T.A. (1992) Science 256, 1783–1788.
[10] Ollis, D.L., Brick, P., Hamlin, R., Xuong, N.G. and Steitz, T.A. (1985) Nature 313, 762–766.
[11] Mitchell, E.M., Artymiuk, P.J., Rice, D.W. and Willett, P. (1990) J. Mol. Biol. 212, 151–166.
[12] Grindley, H.M., Artymiuk, P.J., Rice, D.W. and Willett, P. (1993) J. Mol. Biol. 229, 707–721.
[13] Artymiuk, P.J., Grindley, H.M., Park, J.E., Rice, D.W. and Willett, P. (1992) FEBS Lett. 303, 48–52.
[14] Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer Jr., E.F., Brice, M.D., Rodgers, J R., Kennard, O., Shimanouchi, M. and Tasumi, M. (1977) J. Mol. Biol. 112, 535–542.
[15] Abola, E.E., Bernstein, F.C., Bryant, S.H., Koetzle, T.F. and Weng. J. (1987) in: Crystallographic Databases: Information Content, Software Systems, Scientific Applications (Allen, F.H., Bergeroff, G. and Sievers, R. eds) pp. 107–132, Data Commission of the International Union of Crystallography, Bonn
[16] Kabsch, W. and Sander, C. (1983) Biopolymers 22, 2577–2637.
[17] Bron, C. and Kerbosch, J. (1973) Commun. ACM 16, 575–577.
[18] Jones, T.A. (1985) Methods Enzymol. 115, 157–171.
[19] Yang, W., Hendrickson, W.A., Crouch, R.J. and Satow, Y. (1990) Science 249, 1398–1405.
[20] Fletterick, R.J., Bates, D. and Steitz, T.A. (1975) Proc. Natl. Acad. Sci. USA 72, 38–42.
[21] Flaherty, K.M., DeLuca-Flaherty, C. and McKay, D.B. (1990) 346, 623–628.
[22] Kabsch, W., Mannherz, H.G., Suck, D., Pai, E. and Holmes, K.C. (1990) Nature 347, 37–44.

[23] Hurely, J.H., Faber, H.R., Worthylake, D., Meadow, N.D., Roseman, S., Pettigrew, D.W. and Remington, S.J. (1993) Science 259, 673–677.
[24] Kraulis, P.J. (1991) J. Appl. Crystalogr. 24, 946–950.
[25] Sternberg, M.J.E. and Thornton, J.M. (1975) J. Mol. Biol 105, 367–382.
[26] Naharro, G., Robbins, K.C. and Reddy, E.P. (1984) Science 223, 63–66.
[27] see Blow, D.M. (1990) Nature 333, 694–695.
[28] see Creighton, T.E. (1984) Proteins: Structures and Molecular Properties, pp. 263–264, Freeman, New York.
[29] Matthews, B.W. and Rossmann, M.G. (1985) Methods Enzymol. 115, 397–420.
[30] Horiuchi, T., Maki, H. and Sekiguchi, M. (1984) Mol. Gen. Genet. 195, 17–22.
[31] Pavletich, N.P. and Pabo, C.O. (1991) Science 252, 809–817.
[32] Ben-Artzi, H., Zeelon, E., Gorecki, M. and Panet, E. (1992) Proc. Natl. Acad. Sci. USA 89, 927–931.
[33] Hostomsky, Z., Hudson, G.O., Rahmati, S. and Hostomska, Z. (1992) Nucleic Acids Res. 20, 5819–5824.
[34] Peliska, J.A. and Benkovic, S.J. (1992) Science 258, 1112–1118.