ELSEVIER

Development of computational and graphical tools for analysis of movement and flexibility in large molecules

Paul A. Keller, Scott P. Leach, Tien T.T. Luu, Stephen J. Titmuss, and Renate Griffith

Department of Chemistry, University of Wollongong, Wollongong, Australia

We developed a computer program for the calculation and display of the difference distance matrices (DDMs) of macromolecules that has the ability to compare multiple structures simultaneously. To demonstrate its use, a data set of atoms for superimposition of the HIV-1 reverse transcriptase enzyme was defined using the coordinates for the 21 available crystal structures of this enzyme and its complexes. The DDM technique for superimposition data set generation allows selection of atoms that are invariant in all structures, is free from user bias, and represents the most accurate and precise method of producing such subsets. Comparison of this technique was made against other published methods of generating superimposition data sets, and it was found that significant differences in magnitude and trends of atom movements are observed depending on which data set was used.

Keywords: superimposition, difference distance matrix, HIV-1 reverse transcriptase, protein flexibility

INTRODUCTION

A major focus of computer-aided molecular modeling of large biomolecules has been the study of rigid systems, or simple conformational changes. Such studies have yielded much information and powerful insights into these systems. However, the dynamic function of proteins requires analyses into the movement and flexibility of these molecules. To this end, we have started to use difference distance matrices (DDMs) as the basis for a technique in these types of analyses. In an initial study, a suitable superimposition data set was required for investigation into the dynamic function of the known flexible enzyme HIV-1 reverse transcriptase (RT). There are a number of reasons for choosing this well-studied protein. It exhibits considerable flexibility in both the action of inhibitors and in its varied catalytic functions. Further, there are numerous different x-ray structures available for analysis, allowing for good variation in complexed inhibitors and other parameters. Through the use of DDM analysis, a data set of C_α atoms was generated, over a range of residues and encompassing most regions of the enzyme, suitable for superimposition studies.

Color Plates for this article are on page 299.

Corresponding author: P.A. Keller, Department of Chemistry, University of Wollongong, Wollongong, Australia, 2522. Tel.: +61 2 4221 4692; fax: +61 2 4221 4287.

E-mail address: paul_keller@uow.edu.au (P.A. Keller)

Difference distance matrices

The generation of DDMs is well documented.[1] Briefly, they are constructed by calculating the distances r_{ij} between the i^{th} and j^{th} amino acids, i.e., every combination of two residues in a single protein. In our case, the coordinates of the backbone C_α atoms are used to represent the position of an amino acid. These distances are stored in an $n \times n$ square matrix with the residue numbers i and j along the axes, where n is the number of residues in the protein. This internally referenced distance matrix (DM) contains all the structural information required to regenerate the relative Cartesian coordinates of the molecule, with the exception of chirality. Pairs of residues that are close to each other in the three-dimensional structure will have small values of r_{ij}, while those that are distant, even if they are close together in the polypeptide sequence, will have larger r_{ij} values.

A DDM comparing two protein conformations A and B is constructed by first calculating separate DMs containing the distances r_{ij}^{A} and r_{ij}^{B} from conformations A and B, respectively. Subtraction of one DM from the other, by calculating the difference Δr_{ij} between elements r_{ij}^{A} and r_{ij}^{B} that have

Journal of Molecular Graphics and Modelling 18, 235–241, 2000

655 Avenue of the Americas, New York, NY 10010

1093-3263/00/$–see front matter
PII S1093-3263(00)00028-0

corresponding values of *i* and *j*, yields the DDM comparing the conformations of A and B. This is constructed more directly by calculating:

$$\Delta r_{ij} = |r_{ij}^{A} - r_{ij}^{B}|$$

for all combinations of *i* and *j* and storing in a square matrix with *i* and *j* along the axes.

If the two conformations A and B are identical, even if not superimposed, the corresponding distances r_{ij}^{A} and r_{ij}^{B} will be identical, and so the difference Δr_{ij} is equal to zero. However, if a residue *x* is in a different position in one structure it will have different distances to every other residue *j* in the protein and the differences in the DDM will be nonzero for every Δr_{ij} value where $i = x$. Thus, regions of dissimilarity between the two structures will appear as clusters of nonzero difference values in the DDM, with the value of Δr_{ij} indicating the magnitude of conformational change.

DDMs represent scalar movements between structures and do not give any information on the direction of the movement. The information gained can be used in two ways: either in direct analyses of movement, or in the creation of subsets for subsequent experimentation, e.g., the selection of atoms that behave consistently, for the generation of subsets for superimposition studies. Examples of the use of DDMs in protein chemistry include the analysis of human cardiac troponin C, complementing nuclear magnetic resonance structural studies by comparing the relative orientations of helices between Ca^{2+}-bound species and the apo form.[2] Other examples include the identification of rigid domains in human hemoglobin,[3] the extraction of geometrically similar substructures,[4] and the definition of the nature of the response of calbindin D_{9k} to ion binding.[5] The use of DDMs in the analysis of HIV-1 RT has been confined to identification of rigid domain regions,[6] concluding that RT can undergo a specific swiveling motion, probably during the process of elongation of the nucleic acid duplex.

Superimposition studies

The generation of subsets for superimposition can take several forms. A traditional method is to perform best-fit superimpositions of all residues (usually all C_α atoms), or a subset of residues followed by an analysis of the regions that differ. However, the shortcomings of this latter methodology are well documented; it is vulnerable to misinterpretations and error due to user bias in choice of atoms to superimpose.[5,7] Further, the choice of different sets of atoms for superimposition can give rise to different results.[8] This method is not validated by the use of root mean square distance (RMSD) optimization, as variations in results can still be obtained depending on the choice of atoms for superimposition.[5] Further, RMSD only provides information on the absolute deviation of atomic coordinates in the different superimposed molecules and does not provide any directionality of the structural changes.[5] The use of rigid-body superimpositions can lead to discrepancies due to differences in the orientation and position of secondary structure elements[7] within the proteins, and by the subsequent minimization of RMS distances.[9]

Additionally, there are different schools of thought as to which are the best regions to choose for a superimposition data set. One method is to select atoms that are directly surrounding the region to be examined, either by choosing a sequence of residues[10] or by creating a subset of residues within a given radius of the region to be examined.[11] Other methods involve the inclusion of two regions at either end of the molecule to be examined[6] or the determination of an invariant core by a "comparison of superimpositions between three structures based on individual domains," giving rise to a "conserved core region of the enzyme."[12]

The use of all residues for superimposition also is error prone and is likely to induce an averaged-out effect over the entire protein.[6,7] This would tend to obscure any systematic alterations and further obliterate any small changes in movement that might exist.

The use of DDMs in determining regions of biomolecules that do and do not move has been documented and the advantages explained. These include the objective determination of structural differences based on interatomic distances,[5,6] with these changes being independent of magnitude, and the ready identification of regions and domains that move (or do not move), both independently or as rigid bodies.[6] Most significantly, the use of DDMs is free from user bias.[7]

In order to develop our techniques for the investigation and analysis of movement and flexibility in large molecules, we developed our own DDM program, incorporating graphical outputs for the visualization of data. We used this program to establish a data set of C_α atoms within the HIV-1 RT enzyme that do not move on binding of an inhibitor or an enzyme substrate, for the 21 x-ray crystallographic files available at the time of preparation. We report the results from our studies in the preparation of this data set for superimposition.

RESULTS AND DISCUSSION

ProFlex: A complete DDM computer program

ProFlex is a Java 1.2 compliant computer program, making it platform independent. It is is menu driven through a graphical interface. It generates DDMs through the importation of coordinate data, followed by the stripping of residues (atoms) that are not common to all molecules to be examined. The program has no limit to the number of proteins that it can analyze concurrently, and it is not limited by the number of individual chains present within each structure. It has the option of performing calculations on different types of atoms (C_α only, side chains only, all atoms), and the limitation of the number of atoms to be analyzed is defined by hardware space and not the software itself. It is able to generate subsets based on user-defined parameters of distance moved. Output consists of contour graphs, with the third dimension illustrating the distance moved through user-defined color coding. The resolution and accuracy of these outputs is sufficient such that zooming-in to examine individual and specific residues is possible. Individual residues can be identified by placing the mouse cursor over the relevant section of the contour graph, and the program will automatically display which residue is being examined. Alternatively, the color-coded distance moved subsets can be superimposed onto the backbone of the enzyme and examined by rotating the enzyme within the graphical interface, allowing for identification of residues that are remote in sequence but close in space. All of the outputs can be exported as either GIF or JPEG files. It is proposed in the near future to incorporate into the program the ability to superimpose structures based on

DDM outputs and to generate graphs of interatomic distances from these superimpositions.

Proflex can automatically analyze multiple structures by comparing final DDM outcomes for each structure. This can be achieved by loading all structures to be analyzed within the same calculation, which will produce a composite matrix and output. Conserved regions can be identified based on the user-defined color-coded distance parameters (for example, 0–2 Å). Regions of similar or conserved movement that are close in space, but remote in sequence, can be detected using the Proflex function of mapping the user-defined regions onto the protein backbone using the same color coding.

A typical example of the output from this program is illustrated in Color Plate 1, which shows the DDM of the HIV-1 RT and a complex with the inhibitor nevirapine bound. Shown is a contour graph (A), with a region zoomed in to illustrate finer detail around the fingers region of this enzyme (B), and a further zoom to illustrate individual residues (C). Regions of blue correspond to movement of 0–2 Å, which we define as regions that do not move, within the confines of the resolutions of the x-ray data imput. These regions are easily identified from these contour graphs by analyzing vertical "streaks" with the color-coding classifying regions/residues that possess similar magnitudes of distance moved. It should be emphasized that an unbiased analysis of a residues movement is best achieved by comparison against all other residues within the protein structure. Therefore, when zooming-in on regions for greater precision in analysis (Color Plate 1B and C), the vertical axis should remain the same as this represents the internal cartesian coordinates of the rest of the protein for the comparison against the individual residue or atom under investigation. Reducing the vertical scale will result in an incomplete analysis, likely to give rise to a biased result.

HIV-1 RT inhibition and associated conformational changes

The DDM approach has been used to characterize domain movements in RT and to identify regions of structural rigidity that may be suitable as the basis for further superimposition studies.[6] Other studies also showed that conformational changes of the RT heterodimer are restricted to selected regions of the enzyme. The versatility of the ProFlex program allows for analysis of the flexibility and movement of multiple structures[12]; however, similar (albeit smaller) analyses of the HIV-1 RT enzyme already are well documented.[13]

Data set generation

The ability of the ProFlex program to directly compare multiple sets of data allows for a concurrent analysis of all available x-ray crystal data for a given structure. The x-ray crystal structures used in these studies were downloaded from the PDB and are listed in Table 1, together with their resolutions, crystal packing group, completeness, and inhibition values where reported.

Table 1. X-Ray crystallographic structures of HIV-1 RT used in this study

PDB file	HIV-1 RT structure	Crystal form	Resolution range (Å)	Data completeness	Reference‡	IC^{50} (nM)§
1HMV	Unliganded RT	C2	6.0–3.2	85	11	—
1HNI	RT/Br-α-APA	C2	10.0–2.8	78.5	14	—
1HNV	RT/8-C1 TIBO (R86183)	C2	10.0–3.0	81	15,16	4.6
1KLM*	RT/BHAP (U-90152)	$P2_12_12_1$	20.0–2.65	86.8	17,18	1100/260
1REV*	RT/9-C1 TIBO (R82913)	$P2_12_12_1$	25.0–2.6	80.7	19,20	600/1.5
1RT1	RT/MKC-422	$P2_12_12_1$	25.0–2.55	95.6	21	8
1RT2	RT/TNK-651	$P2_12_12_1$	20.0–2.55	85.3	21	6
RT3	RT/1051U91 RTMC mutant	$P2_12_12_1$	25.0–3.0	83.3	22	400
1RTH	RT/1051U91	$P2_12_12_1$	25.0–2.2	81.4	23,24	400
1RTI	RT/HEPT	$P2_12_12_1$	25.0–3.0	86.3	23,25	17000
1RTJ*	Unliganded RT	$P2_12_12_1$	25.0–3.5	89.5	10	—
1TVR	RT/9-C1-TIBO (R82913)	C2	10.0–3.0	84#	15	33
1UWB	RT/8-C1 TIBO Tyr181Cys mutant	C2	15.0–3.2	85	15	130
1VRT	RT/nevirapine	$P2_12_12_1$	25.0–2.2	87.1	23,24	84
1VRU	RT/C1-α-APA (R90385)	$P2_12_12_1$	25.0–2.4	86.5	23,26	5
1RTD	TR/DNA/dNTP	$P2_12_12_1$	12.0–3.2	92.2	27	—
2HMI	RT/DNA/Fab	$P3_212$	8.0–2.8	74.8	28	—
3HVT	RT/nevirapine	C2	8.0–2.9	95.6	29	84
1DLO	Unliganded RT	C2	8.0–2.7	99.5#	30	—
BQM	RT-HBY097	C2	10.0–3.1	99.1	31	45
BQN	RT-HBY 097 Tyr188Leu mutant	C2	10.0–3.1	95.1	31	600

* Crystal structures were prepared by soaking out HEPT from a HEPT/RT complex, and then soaking in the new inhibitor, if applicable.
§ IC_{50} values used were obtained preferably from the same publication that reported the crystallographic data. In cases where this was not possible, IC_{50} were obtained from publications reporting the synthesis of the inhibitors.
‡ First (and second reference for 1hnv.pdb) refers to the crystal structure and may include the IC_{50} value. Following references are for IC_{50} data.
# Data completeness overall for range.

Table 2. Threshold determination of the 0–2 Å subset of HIV-1 reverse transcriptase crystal structures in DDMs

PDB file	Percent threshold	Residues aligned
BQM	64.70	971
BQN	63.97	971
HNI	67.30	971
HNV	63.01	971
KLM	73.72	936
REV	70.08	935
RT1	71.30	940
RT2	73.99	939
RT3	73.74	917
RTH	65.15	952
RTI	70.16	946
RTJ	72.24	956
TVR	64.30	971
UWB	64.63	971
VRT	68.73	925
VRU	65.51	935
RTD	81.03	966
HMI	79.75	971
HVT	60.76	945

All comparisons are against the unliganded 1DLO.pdb structure.

Analysis of many crystal structures of HIV-1 RT allows the generation of a data set of residues whose C_α atoms do not move on binding of a non-nucleoside inhibitor, or the binding of an enzyme substrate (dsDNA, dNTP). The initial step in this data set generation was the extraction of residues whose C_α atoms move by <2 Å based on DDM analysis. This value can be varied, but our choice is an estimation based on the resolutions of the crystal structures used. For this data set, lowering this limit would have the effect of increasing the error of the output information and increasing the limit will simply lower the accuracy of the results.

When analyzing a DDM, it is necessary to determine whether individual atoms fall into the classes of distance moved (e.g., 0–2 Å). Therefore, it is necessary to determine a threshold for the number of points needed within a range for each atom to be included in a particular subset. The ProFlex program has the option for a user-defined percentage threshold to be implemented; alternatively, the program can automatically determine this threshold. This automatically determined threshold is based on the total number of points within the entire DDM that are within the subset range to be analyzed (e.g., 0–2 Å) as a percentage of the total number of points within the DDM matrix. The application of this automated threshold determination in the generation of our superimposition subset calculated a percentage threshold for each DDM comparison. This is listed in Table 2, along with the total number of residues aligned for each.

The final data set of residues that do not move on binding of non-nucleoside inhibitors is shown in Figure 1. This analysis is based on movements induced not only by DNA bound in the enzyme, but also by numerous non-nucleoside inhibitors, including nevirapine, 1051U91, Br-α-APA, Cl-α-APA, 8-Cl-TIBO, 9-Cl-TIBO, HEPT, MKC-422, TNK-651, and BHAP. Also included in the analysis is a DNA/Fab structure and another with 8-Cl-TIBO bound to a Tyr181Cys mutant. The features of these x-ray structures are summarized in Table 1. The DLO crystal structure was chosen as the unliganded RT crystal structure over 1HMV and 1RTJ due to its higher resolution and data completeness, and because of the known irregularities in the crystal collection method of 1RTJ.[12] This subset represents a total of 376 residues (41% of the total aligned residues) encompassing significant portions of the protein (Figure 1 and Color Plate 2). Color Plate 2 illustrates the distribution of the residues in this subset (red) on the backbone of the enzyme. The regions that are not included in the subset include the fingers, the thumb, and some of the RNaseH domain.

The value of this methodology for the superimposition data set generation can be illustrated by directly comparing outcomes of superimposition using different reported techniques for the generation of such data sets. This is illustrated in Figure 2, where the displacement of C_α atoms in the region of the polymerase active site of HIV-1 RT on binding of an inhibitor (Cl-α-APA) is plotted using different reported superimposition data set generation methods. Our DDM technique is plotted in black and consists of the superimposition of 1528 pairs of atoms (backbone atoms of 376 residues), yielding an RMS of 1.89. Method 1 is based on a subset defined for HIV-1 RT[10] using the C_α atoms from the β4, β7, and β8 strands, which have been defined as moving as rigid bodies on forming of the RT-NNI complexes. It consists of 17 pairs of C_α atoms and yields an RMS of 0.57. Method 2 is based on a reported

p66: residues: 4-6, 95-107, 162-163, 180-181, 188-200, 202-205, 226, 234-235, 237-239, 317, 319, 323, 339-345, 349-353, 365-366, 368-402, 405-419, 428-436, 439, 493, 530.

p51: residues: 6-7, 18-45, 54-64, 71-84, 97-111, 113-117, 121, 123-138, 140-174, 176-184, 186-192, 197-198, 201-202, 208, 252, 254-264, 267, 274, 277, 280-282, 284, 296, 298-300, 303-307, 320, 322, 329, 331, 333-355, 364-393, 397-417.

Figure 1. Data set of C_α atoms of residues of HIV-1 reverse transcriptase that do not move on binding of non-nucleoside inhibitors or substrates. This data set is independent of the type of ligand bound.

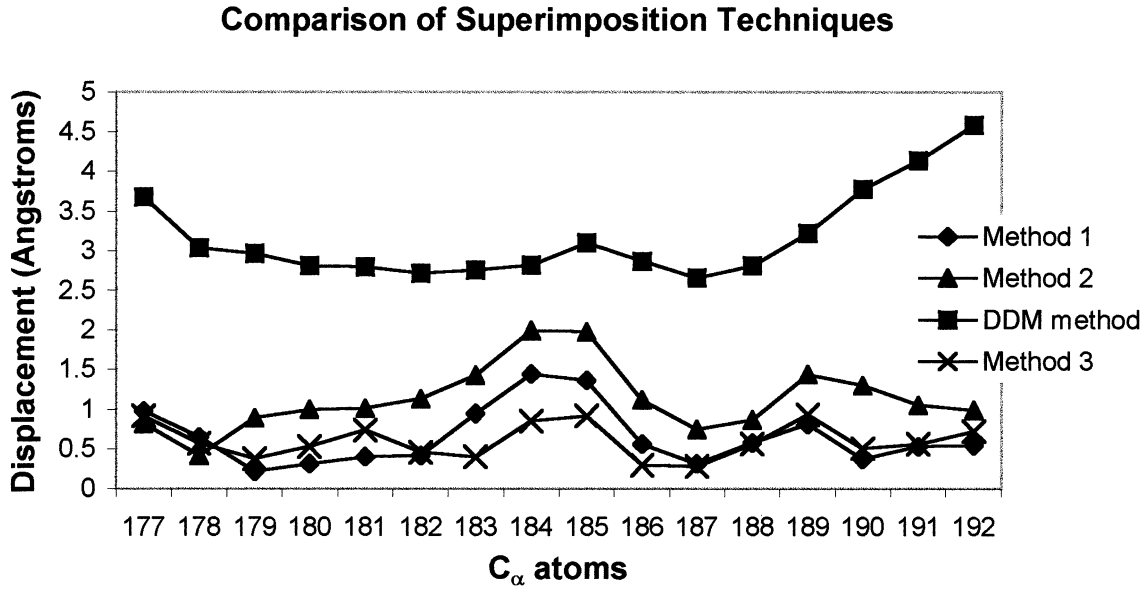


Figure 2. Comparison of superimposition techniques (see text for details on methods).

procedure[10] and is a superimposition of 108 pairs surrounding the polymerase active site and including all C_α atoms within this defined boundary. This corresponds to a radius of 13.3 Å and yields an RMS of 2.5. Method 3 is a superimposition based on a data set of 16 pairs of atoms, as reported by Ding et al[28] and is based on a sequence of C_α atoms that includes the site to be examined. Figure 2 clearly shows the washing out effects arising from superimposition data sets that consist of atoms that are too tightly focused on the area of study. From these techniques, movements of the C_α atoms of the PAS and surrounding residues have been defined mostly as less than 1.5 Å. With our DDM technique, which superimposes all residues in the enzyme that do not change on ligand binding, movements between 3 and 4.5 Å have been shown to occur. Our RMS of 1.89 can be stated as being exceptionally good when compared to the RMS values of the other techniques (0.57, 2.50, and 0.64), considering the size of the different data sets (1528 pairs of atoms compared to 17, 108, and 16). A further aspect illustrated by this graph is the possibility of not only reading an incorrect magnitude of movement, but the increased possibility of an incorrect interpretation of the trends that might be occurring. This is illustrated in Figure 2, which shows residues 177–188 following the same movement trends in all cases, despite some differences in the magnitude of the change. However, residues 189–192 show alarming disparities with our DDM methodology, demonstrating a trend of increasing distances moved, whereas the other methods show a trend of decreasing movement. The relative magnitude of the two main movements (around residues 184 and 189) also are misrepresented by methods 1–3, with an overemphasis on the regions under study.

CONCLUSIONS

We developed a program for the calculation of the DDMs of macromolecules that has the ability to compare multiple structures simultaneously. This is coupled with the ability to analyze data in a variety of ways through the use of an built-in graphing package. We utilized these functions to initiate a study that compares and analyzes a large range of available crystal structures of the HIV-1 RT protein. We used our program to in the creation of a data set of residues whose C_α atoms do not move on binding of ligands. This data set now can be used as a basis for superimposition studies, which we have commenced with a detailed study of the polymerase active site of the HIV-1 RT enzyme.[12] Here, we have been able to assess the changes on binding of non-nucleoside inhibitors objectively and quantitatively and compare them with the changes observed on substrate binding.

ACKNOWLEDGMENTS

We would like to thank the Biomolecular Research Centre, University of Wollongong, for financial support. TTTL thanks the University of Wollongong for a postgraduate scholarship. Researchers interested in gaining access to the program are asked to initially contact the authors directly.

REFERENCES

1 Nishikawa, K., Ooi, T., Isoga, Y., and Saito, N. Tertiary structure of proteins I: Representations and computation of the conformation. *J. Phys. Soc. Jpn.* 1992, **32,** 1331–1337

2 Paeaekkoenen, K., Annila, A., Sorsa, T., Pollesello, P., Tilgmann, C., Kilpelaeinen, I., Karisola, P., Ulmanen, I., and Drakenberg, T. Solution structure and main chain dynamics of the regulartory domain (residues 1–91) of human cardiac troponin C. *J. Biol. Chem.* 1998, **25,** 15633–15638

3 Nichols, W., Rose, G.D., Ten Eyck, L.F., and Zimm, B.H. Rigid domains in proteins: An algorithmic approach to their identification. *Proteins* 1995, **23,** 38–45

4 Lesk, A.M. Extraction of geometrically similar subsstructures: Least-squares and Chebyshev fitting and

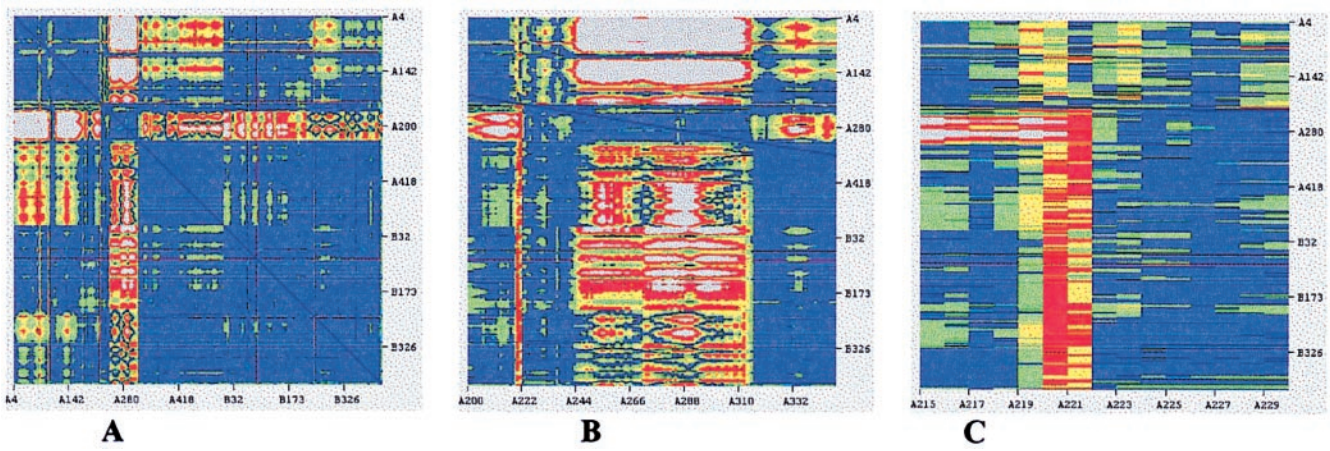
the difference distance matrix. *Proteins* 1998, **33,** 320–328
5 Akke, M., Forsen, S., and Chazin, W.J. Solution structure of $(Cd^{2+})_1$-calbindin D_{9k} reveals details of the stepwise structural changes along the Apo → $(Ca^{2+})_1^{II}$ → $(Ca^{2+})_1^{I,II}$ binding pathway. *J. Mol. Biol.* 1995, **252,** 102–121
6 Jaeger, J., Smerdon, S.J., Wang, J., Boisvert, D.C., and Steitz, T.A. Comparison of three different crystal forms shows HIV-1 reverse transcriptase displays an internal swivel motion. *Structure* 1994, **2,** 869–876
7 Nelson, M.R., and Chazin, W.J. An interaction-based analysis of calcium-induced conformational changes in Ca^{2+} sensor proteins. *Prot. Sci.* 1998, **7,** 270–282
8 May, A.C.W. Pairwise iterative superimpositions of distantly related protein and assessment of the significance of 3-D structural similarity. *Prot. Eng.* 1996, **9,** 1093–1101
9 May, A.C.W., and Johnson, M.S. Improved genetic algorithm-based protein structure comparisons: Pairwise and multiple superimpositions. *Prot. Eng.* 1995, **8,** 873–882
10 Esnouf, R.M., Ren, J., Ross, C., Jones, Y., Stammers, D., and Stuart, D. Mechanism of inhibition of HIV-1 reverse transcriptase by non-nucleoside inhibitors. *Nat. Stuct. Biol.* 1995, **2,** 303–308
11 Rodgers, D.W., Gamblin, S.J., Harris, B.A., Ray, S., Culp, J.S., Hellmig, B., Woolf, D.J., Debouck, C., and Harrison, S.C. The structure of unliganded reverse transcriptase from the human immunodeficiency virus type 1. *Proc. Natl. Acad. Sci. U.S.A.* 1995, **92,** 1222–1226
12 Luu, T.T.T., Keller, P.A., Griffith, R., Titmuss, S.J., and Leach, S.P. Conformational changes in HIV-1 reverse transcriptase induced by non-nucleoside inhibitor and substrate. *Prot. Sci.* (submitted)
13 See references in Table 1
14 Ding, J., Das, K., Tantillo, C., Zhang, W., Clark A.D. Jr., Jessen, S., Lu, X., Hsiou, Y., Jacobo-Molina, A., Andries, K., Pauwels, R., Moereels, H., Koymans, L., Janssen, P.A.J., Smith, R.H. Jr., Koepke, M.K., Michejda, C.J., Hughes, S.H., and Arnold, E. Structure of HIV-1 reverse transcriptase in a complex with the non-nucleoside inhibitor α-APA R95845 at 2.8 Å resolution. *Structure* 1995, **3,** 365–379
15 Das, K., Ding, J., Hsiou, Y., Clark, A.D. Jr., Moereels, H., Koymans, L., Andries, K., Pauwels, R., Janssen, P.A.J., Boyer, P.L., Clark, P., Smith, R.H. Jr., Smith, M.B.K., Michejda, C.J., Hughes, S.H., and Arnold, E. Crystal structure of 8-Cl and 9-Cl TIBO complexed with wild-type HIV-1 RT and 8-Cl TIBO complexed with the Tyr181cys HIV-1 RT drug resistant mutant. *J. Mol. Biol.* 1996, **264,** 1085–1100
16 Ding, J., Das, K., Moereels, H., Koymans, L., Andries, K., Janssen, P.A.J., Hughes, S.H., and Arnold, E. Structure of HIV-1 RT/TIBO R86183 complex reveals similarity in the binding of diverse nonnucleoside inhibitors. *Nat. Struct. Biol.* 1995, **2,** 407–415.
17 Esnouf, R.M., Ren, J., Hopkins, A.L., Ross, C.K., Jones, E.Y., Stammers, D.K., and Stuart, D.I. Unique features of the structure of the complex between HIV-1 reverse transcriptase and the bis(heteroaryl)piperazine (BHAP) U-90152 explain resistance mutations for this non-nucleoside inhibitor. *Proc. Natl. Acad. Sci. U.S.A.* 1997, **94,** 3984–3989
18 Baba, M., Balzarini, J., Pauwels, R., and De Clerq, E. HIV-1 specific reverse transcriptase inhibitors. In: *Anti-AIDS Drug Development Challenges, Strategies and Prospects* (Mohan, P., and Baba, M., eds.). Harwood Academic, Switzerland, 1995, pp. 239–267
19 Ren, J., Esnouf, R., Hopkins, A., Ross, C., Jones, Y., Stammers, D., and Stuart, D. The structure of HIV-1 reverse transcriptase complexed with 9-chloro-TIBO: Lessons for inhibitor design. *Structure* 1995, **3,** 915–926
20 Pauwels, R., Andries, K., Desmyter, J., Schols, D., Kukla, M.J., Breslin, H.J., Raeymaeckers, A., Gelder, J.V., Woestenborghs, R., Heykants, J., Schellekens, K., Janssen, M.A.C., De Clerq, E., and Janssen, P.A.J. Potent and selective inhibition of HIV-1 replication *in vitro* by a novel series of TIBO derivatives. *Nature* 1990, **343,** 470–474
21 Hopkins, A.L., Ren, J., Esnouf, R.M., Willcox, B.E., Jones, E.Y., Ross, C., Miyasaka, T., Walker, R.T., Tanaka, H., Stammers, D.K., and Stuart, D.I. Complexes of HIV-1 reverse transcriptase with inhibitors of the HEPT series reveal conformational changes relevant to the design of potent non-nucleoside inhibitors. *J. Med. Chem.* 1996, **39,** 1589–1600
22 Ren, J., Esnouf, R.M., Hopkins, A.L., Jones, E.Y., Kirby, I., Keeling, J., Ross, C.K., Larder, B.A., Stuart, D.I., and Stammers, D.K. 3′-Azido-3′-deoxythymidine drug resistance mutations in HIV-1 reverse transcriptase can induce long range conformational changes. *Proc. Natl. Acad. Sci. U.S.A.* 1998, **95,** 9518–9523
23 Ren, J., Esnouf, R., Garman, E., Somers, D., Ross, C., Kirby, I., Keeling, J., Darby, G., Jones, Y., Stuart, D., and Stammers, D. High resolution structures of HIV-1 RT from four RT-inhibitor complexes. *Nat. Struct. Biol.* 1995, **2,** 293–302
24 Hargrave, K.D., Proudfoot, J.R., Grozinger, K.G., Cullen, E., Kapadia, S.R., Patel, U.R., Fuchs, V.U., Mauldin, S.C., Vitous, J., Behnke, M.L., Klunder, J.M., Pal, K., Skiles, J.W., McNeil, D.W., Rose, J.M., Chow, G.C., Skoog, M.T., Wu, J.C., Schmidt, G., Engel, W.W., Eberlein, W.G., Saboe, T.D., Campbell, S.J., Rosenthal, A.S., and Adams, J. Novel non-nucleoside inhibitors of HIV-1 reverse transcriptase. 1. Tricyclic pyridobenzo- and dipyridodiazepinones. *J. Med. Chem.* 1991, **34,** 2231–2241
25 Miyasaka, T., Tanaka, H., Baba, M., Hayakawa, H., Walker, R.T., Balzarini, J., and De Clerq, E. A novel lead for specific anti-HIV agents: 1-[(2-hydroxyethoxy)methyl]-6-(phenylthio)thymine. *J. Med. Chem.* 1989, **32,** 2507–2509
26 Pauwels, R., Andries, K., Debyser, Z., van Dael, P., Schols, D., Stoffels, P., DeVreese, K., Woestenborghs, R., Vandamme, A.M., Janssen, C.G.M., Anne, J., Cauwenbergh, G., Desmyter, J., Heykants, J., Janssen, M.A.C., De Clerq, E., and Janssen, P.A.J. Potent and highly selective human immunodeficiency virus type 1 (HIV-1) inhibition by a series of α-anilinophenylacetamide derivatives targeted at HIV-1 reverse transcriptase. *Proc. Natl. Acad. Sci. U.S.A.* 1993, **90,** 1711–1715
27 Huang, H., Chopra, R., Verdine, G.L., and Harrison, S.C. Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: Implications for drug resistance. *Science* 1998, **282,** 1669–1675
28 Ding, J., Das, K., Hsiou, Y., Sarafianos, S.G., Clark,

A.D. Jr., Jacobo-Molina, A., Tantillo, C., Hughes, S.H., and Arnold, E. Structure and functional implications of the polymerase active site region in a complex of HIV-1 RT with a double-stranded DNA template-primer and an antibody Fab fragment at 2.8 Å resolution. *J. Mol. Biol.* 1998, **284,** 1095–111

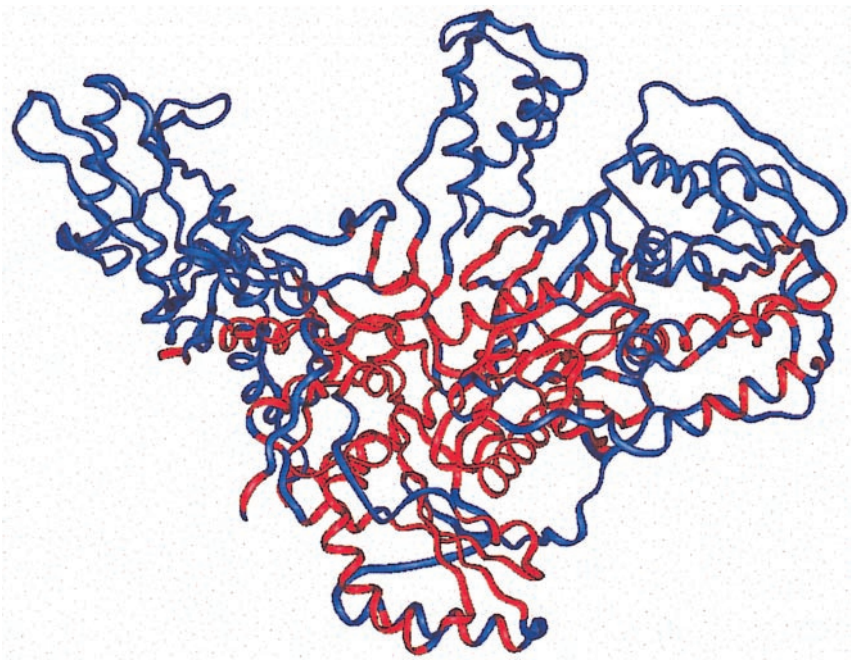
29. Smerdon, S.J., Jäger, J., Wang, J., Kohlstaedt, L.A., Chirino, A.J., Friedman, J.M., Rice, P.A., and Steitz, T.A. Structure of the binding site for non-nucleoside inhibitors of the reverse transcriptase of human immunodeficiency virus type 1. *Proc. Natl. Acad. Sci. U.S.A.* 1994, **91,** 3911–3915
30. Hsiou, Y., Ding, J., Das, K., Clark A.D. Jr., Hughes, S.H., and Arnold, E. Structure of unliganded HIV-1 reverse transcriptase at 2.7 Å resolution: Implications of conformational changes for polymerisation and inhibition mechanisms. *Structure* 1996, **4,** 853–860
31. Hsiou, Y., Das, K., Ding, J., Clark, A.D. Jr., Kleim, J.P., Roesner, M., Winkler, I., Riess, G.A., Hughes, S.H., and Arnold, E. Structures of Tyr188Leu mutant and wild-type HIV-1 reverse transcriptase complexed with the non-nucleoside inhibitor HBY 097: Inhibitor flexibility is a useful design feature for reducing drug resistance. *J. Mol. Biol.* 1998, **284,** 313–323

Paul A. Keller, Scott P. Leach, Tien T.T. Luu, Stephen J. Titmuss, and Renata Griffith

Development of computational and graphical tools for analysis of movement and flexibility in large molecules



Color Plate 1. Difference distance matrix for nevirapine-bound HIV-RT versus the native enzyme (A). Contour graph B shows the same, zoomed into the fingers region, allowing for finer analyses of regions of enzyme movement. Contour graph C illustrates the resolution possible by zooming-in to individual residues. Color coding corresponds to the following movements: blue; 0.0–2.0 Å; green, 2.0–4.0 Å; yellow, 4.0–7.0 Å; red, 7.0–11.0 Å; gray, > 11.0 Å; white, no data.



Color Plate 2. Superimposition data set (see Figure 1) overlaid onto the backbone of the reverse transcriptase enzyme. Red signifies that a residue is included in this subset of residues that do not move. Regions of the enzyme that have been shown to be not included are the fingers, the thumb, and some of the RNaseH domain.