Evaluation of Ligand Overlap by Atomic Parameters

Scott A. Wildman and Gordon M. Crippen*

College of Pharmacy, University of Michigan, Ann Arbor, Michigan 48109-1065

Received July 5, 2000

The overlap of ligand atoms has been analyzed for 32 common enzyme systems. The ligand alignment was determined by superposition of the experimentally determined protein structures. Comparison of the overlapping atoms in terms of atomic contribution to partition coefficient and molar refractivity shows that in most cases ligand atoms overlap with atoms of a similar nature, as should be expected. A new statistic, the mean maximal atomic deviation (MeMAD), is determined and validated for each system studied by comparison to randomized data. In almost all cases, the MeMAD is well separated from the randomized outcome. This result indicates the validity of the atom-based physicochemical parameters in 3-D QSAR methods.

INTRODUCTION

For years, medicinal chemists have built new drug molecules by slightly modifying structures known to bind to the desired target. The standard practice has often been to follow a "chemical intuition" regarding suitable replacements for particular atoms or functional groups. More recently the ideas of quantitative structure−activity relationships (QSAR),[1] pharmacophore modeling,[2] three-dimensional QSAR,[3] and related techniques have all contributed to the ability to generate new drug molecules. While many studies have tested the validity of each of these methods, little if any work has been done to test the chemists' intuition.

The determination of ligand overlap, or alignment, by superposition of the protein structures[4] has been used by several groups for QSAR and structure-based molecular design purposes. Marshall and co-workers[5] used structure-based alignment in CoMFA studies several years ago, a practice which is now common when protein structure data exists. Jalaie and Erickson[6] took this approach one step further. When no experimental structure for photosystem II could be determined, they built a homology model based on purple bacterium reaction center and applied the DOCK methodology[7] to locate their ligands in the protein structure, using the resulting structural alignment for CoMFA analysis.

We use a ligand alignment determined by superposition of the experimentally determined protein structures in a study of chemical intuition in terms of the SLOGP and SMR atomic physicochemical parameters of Wildman and Crippen.[8] The mean maximal atomic deviation (MeMAD) is calculated for each of 32 biological systems.

COMPUTATIONAL METHODS

All calculations in this work were completed using the Molecular Operating Environment (MOE).[9] The systems to be studied were identified by Enzyme Commission (E.C.) number from the ENZYME database.[10] PDB structures of each protein−ligand complex were retrieved from RCSB.[11] Ligand structures were checked against the chemical name, and any inconsistencies were corrected if the intended structure could be determined or else removed. Additional checks were done to ensure that each system contained only protein structures from a single species, that those structures were generated from diffraction or NMR data, and that all X-ray determined structures were of sufficient resolution, better than 4.0 Å RMSD. All solvent and nucleoside chains were deleted. Any protein chains less than 75% identical in sequence to the majority of chains identified for the system were removed. This allowed for inclusion of point mutations and sequence gaps in the structure determined, but further verified that all proteins used in a system were from the same species. A list of structures used for each system is available from the authors.

The sequences for each protein were aligned using the standard MOE-Align[9] methodology. This involves a tree-based multiple alignment technique followed by randomized iterative refinement[12] using the default Gonnet[13] scoring matrix. The exact method of sequence alignment is likely not critical due to the very high sequence identity involved. Following sequence alignment, all the structures were simultaneously spatially superimposed using a weighted nonlinear optimization to determine the rigid-body transformations required to maximize the superposition of the corresponding protein atomic coordinates in the alignment. The ligand structures were not factored in the superposition. The RMSD for each superposition (mean over all pairs of proteins of the root-mean-square C_α coordinate deviation) is reported in Table 1 with the E.C. number and the number of PDB structures used for each system included in the study. Once the alignment and superposition were completed, the protein chains were removed to allow for analysis of only the ligands. In systems with bound cofactors present in the structures, the cofactors were treated as part of the ligand. For the ligand molecules, the mean maximal atomic deviation (MeMAD) statistic was calculated.

Two ligand atoms were said to overlap if they were located in different ligands and were separated by less than 1.0 Å in

* Corresponding author phone: (313)763-9722; e-mail: gordon_crippen@um.umich.edu.

10.1021/ci0000880 CCC: $20.00
Published on Web 01/12/2001

Table 1. Systems Studied

name	E.C. no.	RMSD	entries	overlaps	nonzero overlaps
acetylcholinesterase	3.1.1.7	0.3786	13	293	88
alcohol dehydrogenase	1.1.1.1	0.7447	19	1917	846
aldose reductase	1.1.1.21	0.3913	9	515	113
aspartate aminotransferase	2.6.1.1	0.7779	34	1092	580
aspartate transcarbamoylase	2.1.3.2	1.7241	21	894	364
carbonic anhydrase	4.2.1.1	0.2218	101	779	316
chymotrypsin	3.4.21.1	0.7568	27	1283	369
cyclodextrin glycosyltransferase	2.4.1.19	0.3335	12	1059	338
cytochrome P450	1.14.15.1	0.4544	22	1194	185
dihydrofolate reductase	1.5.1.3	0.5732	42	2525	1383
endothiapepsin	3.4.23.22	0.4169	22	1236	557
fructose-1,6-bisphosphatase	3.1.3.11	0.9502	24	1514	787
glutathione s-transferase	2.5.1.18	0.2799	26	1847	1001
glycogen phosphorylase	2.4.1.1	0.8830	29	1813	587
HIV-1 protease	3.4.23.16	0.7166	28	1557	810
HIV-1 reverse transcriptase	2.7.7.49	1.4004	14	286	139
lecithinase A	3.1.1.4	0.6884	6	138	68
lipase	3.1.1.3	0.2690	6	206	41
lysozyme	3.2.1.17	0.5254	16	387	89
micrococcal nuclease	3.1.31.1	0.2461	24	685	337
pancreatic elastase	3.4.21.36	0.3670	28	791	276
papain	3.4.22.2	0.4544	11	149	63
phospholipase C-γ1	3.1.4.11	0.4892	9	249	124
phosphotyrosine phosphatase	3.1.3.48	0.6293	10	260	107
renin	3.4.23.15	0.7430	5	300	88
stromelysin	3.4.24.17	1.4304	15	432	119
thermolysin	3.4.24.27	0.1528	19	513	254
thrombin	3.4.21.5	0.6061	75	6985	1838
thymidine kinase	2.7.1.21	0.5190	10	453	250
thymidylate synthase	2.1.1.45	0.5246	26	2075	1280
trypsin	3.4.21.4	0.6325	47	1107	535
tyrosylprotein kinase	2.7.1.112	0.7473	14	775	291

Euclidian space. If the overlapping atoms did not have equal atomic contributions to octanol/water partition coefficient (logP) and molar refractivity (MR) as determined by the SLOGP method,[8] it was considered a nonzero overlap. Table 1 includes the total number of overlaps for each system as well as the number of nonzero overlaps.

It was found to be the usual situation that each atom in the group of ligands overlapped with one or two atoms in each of several other ligands. For this reason, the error of atomic overlap was taken as the maximum of the absolute deviation for all possible overlaps for a given atom. This quantity was summed over all atoms in all ligand chains in the system and divided by the number of atoms with nonzero overlaps to give the MeMAD as

$$\text{MeMAD} = \frac{1}{n}\sum_{i=1}^{n}\max|p_i - p_j| \;\forall\; i,j \text{ such that } d_{i,j} \leq 1.0 \text{ Å}, |p_i - p_j| \neq 0 \quad (1)$$

where the sum was performed over the set of atoms, i, with nonzero overlaps as defined by $d_{i,j}$, the distance between each atom pair requiring $d_{i,j} \leq 1.0$ Å and $p_{i,j}$, the atomic property values for each atom, requiring $|p_i - p_j| \neq 0$. For comparison purposes, the MeMAD was also computed using randomized data (RAD), where the SLOGP/SMR atomic contributions are shuffled for all nonzero overlaps in the original calculation as

$$\text{RAD} = \frac{1}{n}\sum_{i=1}^{n}\max|(\pi p)_i - (\pi p)_j| \quad (2)$$

as previously defined but with $(\pi p)_{i,j}$ each representing an element from a random permutation of values of the full set of atoms i and j involved in nonzero overlaps (eq 1) for the given system.

RESULTS AND DISCUSSION

This is in no way meant to be a complete survey of RCSB or ENZYME. These systems were chosen for the large number of structures involving different ligands with the same protein from the same species, or the expectation thereof. Several of these systems are commonly reported in computational ligand docking, binding site modeling or 2-D and 3-D QSAR studies as classic examples and standard test cases. It is of note that all systems studied are enzymes that could be identified by an E. C. number and an entry for which could be found in ENZYME. This limitation is a result of difficulties with the searching functions of RCSB, the documentation of which warns "Inconsistencies in the way data are reported within PDB files may lead to unexpected or incomplete results when searching this field.". ENZYME also allowed for simple separation of the structures of a given system by species. As a result, receptors and drug targets without enzymatic activity are not included in the study.

The MeMAD statistic measures the average greatest difference between overlaps in atomic contributions. If the chemists' intuition is, as expected, correct, overlapping atoms should have have differences in atomic contributions to logP and MR of zero or very close to zero. This corresponds to ligands being structurally very similar to one another, which is often the case for tight binding ligands derived through traditional medicinal chemistry methods. Most of the overlaps seen do fall into this category. Large nonzero overlaps, and hence large MeMAD, could be interpreted as questionable

Table 2. MeMAD Calculated by SLOGP and SMR

		RAD			RAD	
name	logP MeMAD	mean	*N*/1000	MR MeMAD	mean	*N*/1000
acetylcholinesterase	0.3895	0.5804	0	1.6236	3.2851	0
alcohol dehydrogenase	0.5733	0.9995	0	2.1286	3.6139	0
aldose reductase	0.7396	0.9096	0	2.3543	3.4635	0
aspartate aminotransferase	0.7961	1.2940	0	2.3710	4.6174	0
aspartate transcarbamoylase	0.7337	1.1755	0	2.9609	3.6851	0
carbonic anhydrase	1.0580	1.2217	0	4.3543	5.5017	0
chymotrypsin	0.4360	0.7391	0	3.1264	4.0083	0
cyclodextrin glycosyltransferase	0.1379	0.1837	0	1.4899	1.7414	0
cytochrome P450	0.2615	0.4202	0	2.4811	2.8401	1
dihydrofolate reductase	0.7551	1.6223	0	2.1464	3.9456	0
endothiapepsin	0.4489	0.7388	0	2.0826	3.1037	0
fructose-1,6-bisphosphatase	0.7379	1.0729	0	2.4646	4.2765	0
glutathione s-transferase	0.5902	1.0357	0	2.7940	4.7748	0
glycogen phosphorylase	0.6579	1.1762	0	2.1867	4.7956	0
HIV-1 protease	0.4495	0.7201	0	1.9771	2.9321	0
HIV-1 reverse transcriptase	0.4767	0.5232	3	1.9736	1.9433	630
lecithinase A	0.3597	0.4557	7	1.2930	2.0980	0
lipase	0.2740	0.3903	9	2.3171	3.1424	3
lysozyme	0.1957	0.4176	0	1.5576	2.0898	0
micrococcal nuclease	0.8086	1.8366	0	2.1573	4.2074	0
pancreatic elastase	0.5662	0.7054	0	2.9269	4.3026	0
papain	0.6008	0.5944	574	2.4245	2.5786	239
phospholipase C-γ1	1.6350	2.1220	0	2.1005	4.7809	0
phosphotyrosine phosphatase	0.6171	1.0880	0	1.6279	4.0758	0
renin	0.2715	0.3967	0	1.4350	2.0676	0
stromelysin	0.3546	0.5923	0	1.5783	2.6038	0
thermolysin	1.2600	2.7141	0	1.3709	4.5019	0
thrombin	0.7009	1.5152	0	2.6988	3.9207	0
thymidine kinase	0.4206	0.7000	0	2.6254	4.5164	0
thymidylate synthase	0.8314	1.1858	0	2.8028	3.9676	0
trypsin	0.6929	1.2141	0	2.1303	3.7803	0
tyrosylprotein kinase	0.5567	0.9448	0	2.4806	3.2003	0

chemical intuition, or as a failure of the SLOGP system for determining atomic contributions to physicochemical properties. However, large MeMAD may also arise in areas of the ligand that do not interact strongly in a discrete manner with the protein, such as linking regions, or may be exposed to solvent. In these areas, atoms would not be required to have similar specific logP or refractivity contributions. On the other hand, near-zero overlaps and the resulting small MeMAD indicate the validity of both the chemists' intuition and the SLOGP method. Small MeMAD would also suggest that such overlaps of atomic properties at specific locations, or specific distances, can be a useful tool for predicting tight-binding ligands for a given system, an indication of the usefulness of atomic property measurements for 3-D QSAR application.

The MeMAD avoids contributions from cofactors and bound ions by only adding in contributions from nonzero overlaps and only counting those overlaps for averaging. It provides the largest interpretation of the error by taking the maximum absolute deviation at each atom. For example, if four atoms defined by the SLOGP system as type C23, the substituted carbon in cyclohexanol for instance, (atomic contribution to logP $= +0.5437$) all overlap with each other and with a single atom of type C04, a phenolic carbon, (logP $= -0.2783$), each of the five atoms would account for an additional 0.7488, the maximum absolute difference in each case, to be included in the total atomic overlap error. If this is the only set of nonzero overlaps in the system, this contribution would be summed and divided by the total number of nonzero overlaps in the system, or MeMAD $= 0.7488 = 5 \times 0.7488/5$.

The MeMADs for each system using both the SLOGP and SMR metrics are presented in Table 2. The average MeMADs for all systems studied are 0.6072 and 2.2513 using SLOGP and SMR, respectively. For comparison purposes, the MeMAD was also computed using randomized data, where the SLOGP/SMR atomic contributions are shuffled for all nonzero overlaps in the original calculation. It is expected that a random permutation of atomic parameters would result in nonzero overlaps with differences much greater than the experimentally determined overlaps causing the randomized MeMAD (RAD) to be greater than MeMAD. A RAD value near to or lower than the MeMAD will occur if the nonzero overlaps have large differences in property value before random permutation, or if all nonzero overlaps are of the same value. This situation can be interpreted as poor superposition of the ligands, or inaccurate calculation of the atomic property values.

The RAD was calculated 1000 times for each system and is presented as the mean of the distribution and the number of times, *N* (in 1000 random permutations) that the resulting value is lower than the MeMAD, as RAD $\leq$ MeMAD. The separation between MeMAD and RAD is presented in this manner because the distribution is usually nonGaussian and often bimodal, making an accurate interpretation of standard deviation impractical. This description may be viewed as the likelihood of finding a better arrangement of atoms involved in nonzero overlaps.

Comparison of the MeMAD to the randomized data statistic, as presented in Table 2, provides a clear indication of the validity both of the statistic and the SLOGP approach. In 30 of 32 systems studied, there is less than a 1.0% chance

Table 3. HIV Results by Groups

	logP	RAD		MR	RAD	
group	MeMAD	mean	*N*/1000	MeMAD	mean	*N*/1000
Ren	0.3574	0.4744	0	1.9486	2.4015	5
Ding	0.4282	0.4723	52	1.6155	1.9264	5

of identifying an arrangement of atoms involved in nonzero overlaps with a lower MeMAD than determined experimentally using either the contributions to logP or refractivity. Lecithinase A has a slight chance of recognizing a better set of overlaps using logP, as does lipase using either metric. In each case, this is due in part to the small number of ligands and therefore small number of nonzero overlaps. This situation will allow one overlap with large atomic deviation to overshadow several near-zero overlaps. More interesting however, are the cases where the MeMAD does not follow the trend, specifically papain and HIV reverse transcriptase.

Papain is the only plant protein included and is one of the smaller systems included, consisting of 11 ligands with 149 overlaps, 63 of which are nonzero. Several of these overlaps have atomic deviations still very close to zero, indicating good comparison of different atom types, with only a few atom overlaps contributing the majority of MeMAD. Of these, four are large only when measured using the logP metric, four others are large only in refractivity, and three overlaps are large in both. This may seem inconsistent, however it is not unexpected as logP and refractivity are not well correlated. By visual inspection of the structures, many of the nonzero overlaps occur in areas that do not interact with the protein and should therefore not drastically alter the binding properties of the ligands.

In the case of HIV reverse transcriptase, not only is the MeMAD for the SMR metric easily bettered by the RAD (63% of the time), but the RMSD of superposition is high as well. A review of the PDB structures included in the study shows that two research groups have produced almost all of the structures used. Therefore, the structures were classified into two groups to be analyzed separately as they are likely different isolates or strains of HIV. Ren and co-workers have provided 1KLM, 1REV, 1RT1, 1RT2, 1RT3, 1RTH, 1RTI, and 1RTJ, and Ding and co-workers have provided 1BQM, 1BQN, 1HNI, and 1HVT of the structures used in the original measurement. 3HVT is included in the second set as it is identified as one of the isolates used by Ding. When the two subsets of structures are treated separately, the RMSD of superposition and the MeMAD for SMR as compared to randomized data are both improved, as shown in Table 3. The logP MeMAD has improved to 0.3574 and 0.4282 compared to randomized values of 0.4744 and 0.4723, while for refractivity the MeMAD has improved to 1.9486 and 1.6155 compared to 2.4015 and 1.9264. There remains a 5.2% likelihood of finding a better atom arrangement for the Ding ligands using the SLOGP metric, but only 0.5% using SMR. This likelihood for the Ren subset is 0.5% or less using either metric. It seems clear that this analysis is the correct treatment of reverse transcriptase and that this protein does indeed follow the established pattern of the other systems studied.

In contrast to random permutations of atomic property values, it may be useful to investigate small deviations in the geometry of the ligand overlap provided by the protein superposition. A perturbation in the position of the ligands by some small random translation and/or rotation (with no ligand flexibility) would correspond to reduced resolution of the ligand overlap, as could be expected from low resolution protein structures, or poor protein superposition. For each ligand in each system studied, a random rigid-body translation and rotation on the order of 1.0 Å and 1.0° was performed with the requirement that the resulting random superposition have more than half as many nonzero overlaps as the original superposition. This random superposition of the 28 ligands of pancreatic elastase resulted in a logP MeMAD of 0.6001 and a MR MeMAD of 2.6957. These values are not significantly different than the protein superposition values of 0.5662 and 2.9269 for logP and MR, respectively. However, the difference does show in the RADs, both of which are often lower than the MeMAD. The average logP RAD is 0.58609, with 735 occurrences of a lower RAD than MeMAD, and the average MR RAD is 2.5308, with 951 occurrences of lower RAD than MeMAD. Random superposition MeMAD and RAD values were calculated for each system studied with similar results.

The overlaps that occur in a random superposition are not the experimental overlaps, as determined by protein superposition, but instead involve atoms that may have little or no similarity and therefore may have large differences in atomic property values before the RAD permutation. When the RAD permutation is then performed, it can serve to decrease the property differences, resulting in a RAD lower than the MeMAD. While we acknowledge the random superposition may not be an intelligent one, the MeMAD and RAD statistics also identified it as such, again suggesting that the protein-based ligand superposition is correct.

CONCLUSIONS

As expected, ligand atoms that overlap in a superposition of the protein structures do indeed share similar atomic contributions to physicochemical properties. This may be interpreted as validation of the long-held "chemists' intuition" and also of the atomic logP/MR approach. When atomic properties are combined with the understanding of the spatial or distance requirements inherent in the ligand structures that bind to a given protein, the model of protein−ligand interaction begins to emerge. Clear applications in this direction include the generation of 3-D QSARs based on these atomic properties at specific distances for use in pharmacology, environmental toxicology, and pharmaceutical design.

ACKNOWLEDGMENT

This work was supported by a grant from the National Institutes of Health (GM59097). S.A.W. would like to thank Mark A. Murcko for his helpful suggestions.

REFERENCES AND NOTES

(1) Kubinyi, H. QSAR and 3D QSAR in Drug Design Part 1: Methodology. *Drug Discov. Today* **1997**, *2*, 457−467. Kubinyi, H. QSAR and 3D QSAR in Drug Design Part 2: Applications and Problems. *Drug Discov. Today* **1997**, *2*, 538−546.

(2) Martin, Y. C. In *Designing Bioactive Molecules*; Martin, Y. C., Willett, P., Eds.; American Chemical Society: Washington, DC, 1997; pp 121−148.

(3) Martin, Y. C. 3D QSAR: Current Scope and Limitations. *Perspect. Drug Discov.* **1998**, *12*, 3−23.
(4) Klebe, G. In 3D QSAR in Drug Design. *Theory, Methods and Applications*; Kubinyi, H., Ed.; ESCOM: Leiden, Netherlands, 1993; pp 173−199.
(5) DePriest, S. A.; Mayer, D.; Naylor, C. B.; Marshall, G. R. 3D-QSAR of Angiotensin-Converting Enzyme and Thermolysin Inhibitors: A Comparison of CoMFA Models Based on Deduced and Experimentally Determined Active Site Geometries. *J. Am. Chem. Soc.* **1993**, *115*, 5372−5384. Waller, C. L.; Oprea, T. I.; Giolitti, A.; Marshall, G. R. Three-Dimensional QSAR of Himan Immunodeficiency Virus (I) Protease Inhibitors. 1. A CoMFA Study Employing Experimentally-Determined Alignment Rules. *J. Med. Chem.* **1993**, *36*, 4152−4160.
(6) Jalaie, M.; Erickson, J. A. Homology Model Directed Alignment for Comparative Molecular Field Analysis: Application to Photosystem II Inhibitors. *J. Comput-Aided Mol. Des.* **2000**, *14*, 181−197.
(7) Ewing, T. J. A.; Kuntz, I. D. Critical Evaluation of Search Algorithms for Automated Molecular Docking and Database Screening. *J. Comput. Chem.* **1997**, *18*, 1175−1189.
(8) Wildman, S. A.; Crippen, G. M. Prediction of Physicochemical Parameters by Atomic Contributions. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 868−873.
(9) *Molecular Operating Environment, Version 2000.02*; Chemical Computing Group, Inc.: Montreal, 2000.
(10) Bairoch A. The ENZYME database in 2000. *Nucleic Acids Res.* **2000**, *28*, 304−305.
(11) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242. http://www.rcsb.org/pdb/.
(12) Berger, M. P.; Munson, P. J. A Novel Randomized Iterative Strategy for Aligning Multiple Protein Sequences. *CABIOS.* **1991**, *7*, 479−484.
(13) Gonnet, G. H.; Cohen M. A.; Benner, S. A. Exhaustive Matching of the Entire Protein Sequence Database. *Science* **1992**, *256*, 1433−1145.
(14) Ren, J.; Esnouf, R.; Garman, E.; Somers, D.; Ross, C.; Kirby, I.; Keeling, J.; Darby, G.; Jones, Y.; Stuart, D.; Stammers, D. High-Resolution Structures of HIV-1 RT from Four RT-Inhibitor Complexes. *Nat. Struct. Biol.* **1995**, *2*, 293. Esnouf, R.; Ren, J.; Ross, C.; Jones, Y.; Stammers, D.; Stuart, D. Mechanism of Inhibition of HIV-1 Reverse Transcriptase by Nonnucleoside Inhibitors. *Nat. Struct. Biol.* **1995**, *2*, 303. Ren, J.; Esnouf, R.; Hopkins, A.; Ross, C.; Jones, Y.; Stammers, D.; Stuart, D. The Structure of HIV-1 Reverse Transcriptase Complexed with 9-chloro-TIBO: Lessons for Inhibitor Design. *Structure* **1995**, *3*, 915. 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.; Stuart, D. I. Complexes of HIV-1 Reverse Transcriptase with Inhibitors of the HEPT Series Reveal Conformational Changes Relevant to the Design of Potent Nonnucleoside Inhibitors. *J. Med. Chem.* **1996**, *39*, 1589. Esnouf, R. M.; Ren, J.; Hopkins, A. L.; Ross, C. K.; Jones, E. Y.; Stammers, D. K.; Stuart, D. I. Unique Features in the Structure of the Complex Between HIV-1 Reverse Transcriptase and the bis(heteroaryl)piperazine (BHAP) U-90152 Explain Resistance Mutations for this Nonnucleoside Inhibitor. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 3984. Ren, J.; Esnouf, R. M.; Hopkins, A. L.; Jones, E. Y.; Kirby, I.; Keeling, J.; Ross, C. K.; Larder, B. A.; Stuart, D. I.; 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.
(15) 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.; Koepke, M. K.; Michejda, C. J.; Hughes, S. H.; Arnold, E. Structure of HIV-1 Reverse Transcriptase in a Complex with the Nonnucleoside Inhibitor alpha-APA R 95845 at 2.8 Å Resolution. *Structure* **1995**, *3*, 365. Ding, J.; Das, K.; Moereels, H. Koymans, L. Andries, K. Janssen, P. A. Hughes, S. H.; Arnold, E. Structure of HIV-1 RT/TIBO R 86183 Complex Reveals Similarity in the Binding of Diverse Nonnucleoside Inhibitors. *Nat. Struct. Biol.* **1995**, *2*, 407. Hsiou, Y.; Das, K.; Ding, J.; Clark, A. D., Jr.; Kleim, J. P.; Rosner, M.; Winkler, I.; Riess, G.; Hughes, S. H.; Arnold, E. Structures of Tyr188Leu Mutant and Wild-type HIV-1 Reverse Transcriptase Complexed with the Nonnucleoside Inhibitor HBY 097: Inhibitor Flexibility is a Useful Design Feature for Reducing Drug Resistance. *J. Mol. Biol.* **1998**, *284*, 313.
(16) Wang, J.; Smerdon, S. J.; Jager, J.; Kohlstaedt, L. A.; Rice, P. A.; Friedman, J. M.; Steitz, T. A. Structural Basis of Asymmetry in the Human Immunodeficiency Virus Type 1 Reverse Transcriptase Heterodimer. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 72.

CI0000880