Quantitative Symmetry in Structure−Activity Correlations: The Near C2 Symmetry of Inhibitor/HIV Protease Complexes

ARTICLE *in* JOURNAL OF THE AMERICAN CHEMICAL SOCIETY · APRIL 2000

Impact Factor: 12.11 · DOI: 10.1021/ja993810n

2 AUTHORS, INCLUDING:

David Avnir
Hebrew University of Jerusalem
384 PUBLICATIONS **16,684** CITATIONS

Quantitative Symmetry in Structure−Activity Correlations: The Near C_2 Symmetry of Inhibitor/HIV Protease Complexes

Shahar Keinan and David Avnir*

Contribution from The Institute of Chemistry and The Lise Meitner Minerva Center for Computational Quantum Chemistry, The Hebrew University of Jerusalem, Jerusalem 91904, Israel

Received October 26, 1999. Revised Manuscript Received February 16, 2000

Abstract: We studied the way in which the binding of inhibitors of human immunodeficiency virus (HIV) protease causes the protein to deviate from its original C_2 symmetric structure. We extended to C_2 symmetry our findings that quantitative chirality is a useful, predictive parameter in enzymatic structure−activity correlations (*J. Am. Chem. Soc.* **1998**, *120*, 6152−6159). We provide a quantitative assessment of this deviation, the degree of C_2-ness, $S(C_2)$, by employing the continuous symmetry measures methodology. The data analyzed was for a group of 13 inhibitor/protease complexes, for which the structures and the binding energies are known. $S(C_2)$ was determined for the inhibitors before and after binding, for each pair of amino acids of the protein, and for the whole protein complexes. We were able to track the spreading of the C_2 distortion throughout the protein to various zones. Maps of iso-symmetry value proved to be a powerful descriptive tool for protein structure−deviation visualization. The main findings are the following: (i) For most inhibitors, the active site imposes its C_2 symmetry on the bound inhibitor, rendering it more C_2 symmetric than its free form and confirming the picture of enzymes as mechanical devices. (ii) The binding energy of the inhibitors correlates with this imposed C_2 symmetry change: the smaller the symmetry change, the better the inhibition. (iii) Analysis of the enzyme's mutant strain V82A (raised against the inhibitors) shows that it has "learned" to cope better with an inhibitor by "following" this symmetry/binding energy correlation. (iv) Symmetry deviations occur in the protein upon binding at remote zones from the active site. Despite variations in the details of these deviations for the different HIV protease/inhibitor complexes, the protein as a whole responds to the various inhibitors with a very similar C_2 symmetry change: a global symmetry-well for this protein, has been identified.

Introduction

We report here a quantitative study of the symmetry variations of a protein which undergoes binding with inhibitors. Specifically we selected the C_2 symmetric protease of human immunodeficiency virus (HIV). This study follows a recent report in which we found that the quantitative degree of chirality of substrates correlates with their efficiency of reaction with active sites.[1] The degree of chirality, a global shape descriptor, was determined by the use of the continuous chirality measure (CCM) methodology [2−4] which treats chirality as a continuous structural property (distinct from chiral/not chiral). The generality of the new type of shape-activity correlation was demonstrated on several receptor/substrate systems, including trypsin and acetylcholinesterase with chiral inhibitors, and the D_2-dopamine receptor with chiral agonists. For some of the studied cases the correlation of activity with structure—hidden when classical parameters such as chain length are taken—was revealed only with this novel shape descriptor. The findings of that study are quite nontrivial because chirality is a global shape parameter and not a specific descriptor of the intricate geometry of the drug or of the active site. We have proposed that these results may indicate two different recognition mechanisms: shape recognition and chemical recognition, the first being a low-resolution determination of the overall shape of the substrate and the second being the classical key−lock mechanism.

Since the CCM is a special case of the broader, general case of continuous symmetry measures (CSM),[5,6] it has been of interest to explore the possibility that quantitative symmetry can be used for the study and characterization of enzymatic activity as well. We were motivated not only by the observations of ref 1 but also by the successful identification of several correlations between the symmetry measure and physical/chemical parameters.[7−10] To test the feasibility of this approach, we have selected the C_2-symmetric protease of HIV and its distortion from that symmetry upon binding with inhibitors.[11]

We recall that one of the crucial stages in the HIV life cycle is the protease-mediated transformation from the immature,

* Author for correspondence. E-mail: david@chem.ch.huji.ac.il.

(1) Keinan, S.; Avnir, D. *J. Am. Chem. Soc.* **1998**, *120*, 6152−6159.

(2) Zabrodsky, H.; Peleg, S.; Avnir, D. *J. Am. Chem. Soc.* **1995**, *117*, 462−473.

(3) Gao, D.; Schefzick, S.; Lipkowitz, K. B. *J. Am. Chem. Soc.* **1999**, *121*, 9481−9482.

(4) Lipkowitz, K. B.; Gao, D.; Katzenelson, O. *J. Am. Chem. Soc.* **1999**, *121*, 5559−5564.

(5) (a) Zabrodsky, H.; Peleg, S.; Avnir, D. *J. Am. Chem. Soc.* **1992**, *114*, 7843−7851. (b) Zabrodsky, H.; Peleg, S.; Avnir, D. *J. Am. Chem. Soc.* **1993**, *115*, 8278−8288.

(6) (a) Pinto, Y.; Salomon, Y.; Avnir, D. *J. Math. Chem.*, **1998**, *23*, 13−29 (b) Salomon, Y.; Avnir, D. *J. Comput. Chem.*, **1999**, *20*, 772−780 (c) Salomon, Y.; Avnir, D. *J. Math. Chem.* **1999**, *25*, 295−308.

(7) (a) Avnir, D.; Katzenelson, O.; Keinan, S.; Pinsky, M.; Pinto, Y.; Salomon, Y.; Zabrodsky Hel-Or, H. In *Concepts in Chemistry*; Rouvray, D. H., Ed.; Research Studies Press: Somerset, 1997; Chapter 9, pp 283−324. (b) Avnir, D.; Zabrodsky Hel-Or, H.; Mezey, P. G. In *Encyclopedia of Computational Chemistry*; Schleyer, P. V. S., Ed.; Wiley: Chichester, 1998; Vol. 4, pp 2890−2901.

(8) Buch, V.; Gershgoren, E.; Zabrodsky Hel-Or, H.; Avnir, D. *Chem. Phys. Lett.* **1995**, *247*, 149−153.

(9) Pinto, Y.; Fowler, P. W.; Mitchell, D.; Avnir, D. *J. Phys. Chem. B* **1998**, *102*, 5776−5784.

(10) Pinto, Y.; Zabrodsky Hel-Or, H.; Avnir, D. *J. Chem. Soc., Faraday Trans.* **1996**, *92 (14)*, 2523−2527.

10.1021/ja993810n CCC: $19.00
Published on Web 04/21/2000

nondangerous virion, to the mature, infective virus. HIV protease inhibitors have thus become a major target for anti-AIDS drug design.[12] The HIV protease is a homodimer consisting of two identical strands, each with 99 amino acids, arranged with perfect C_2 symmetry in the apo state.[13] The C_2 symmetry of the protease and of its active site have served as a useful guideline in the design of inhibitors: C_2-symmetric and nearly C_2-symmetric inhibitors have been tested under the assumption of better drug−active site fit, due to the similar shape.[14] Interestingly, upon binding, most inhibitors break the C_2 symmetry of the protease. It is an important, telltale structural change from which much can be learned; quantitative assessment of this symmetry change is therefore desirable. Having in mind that symmetry is a common feature of many proteins, it is both interesting and important to find whether quantitative symmetry can be used as a structural correlant with activity. Since symmetry—unlike bond lengths, bond angles etc.—is a nonspecific global shape parameter, this question actually translates into the deeper one, already asked in ref 1: can a global shape descriptor be used to identify correlations with activity, in a place where specific-geometry interactions govern the activity?[15] Here we show that the findings of ref 1 can be further extended to C_2 symmetry. It is an important extension because C_2 symmetry is perhaps the most studied biochemical symmetry feature,[16] second only to chirality (which is, in fact the *lack* of improper symmetry). Interestingly, the C_2 symmetry found in nature is quite often only *nearly* so: see ref 17 for a recent example and see our study of the near C_2 symmetry of the purple bacteria photosynthetic reaction center.[18]

The rationale, the approach, the practical solutions and the applications of symmetry measurement by the continuous symmetry measure (CSM) methodology were described in our previous papers.[5,6,19] (For reviews, see ref 7, for other uses of the CSM approach, see ref 20, for other methods to evaluate symmetry, see ref 21). The special case of measuring the degree of C_2-ness was also treated in detail.[6b,18] We recall that the measure, $S(C_2)$, evaluates the minimal distance that the vertexes of a structure have to be translated in order to acquire a desired symmetry (eq 1 below), that $S(C_2) = 0$ means that the shape has an exact C_2 symmetry, and that $S(C_2)$ increases with the deviation from perfect symmetry.

Data Sources and C_2-Measurements

We studied a series of 12 HIV-1 protease inhibitors (Figure 1) and their complexes with the protein, following the work of Bardi et al.[22] The X-ray structures of the complexes are available; their PDB[23] names: 1PRO,[24] 1HVJ, 1HVK, 1HVI, 1HVL,[25] 1HPV,[26] 1HPS,[27] 1HIH,[28] 1SBG,[29] 2UPJ,[30] 1GNO[31] and 1HBV[32]).[33] Binding affinities, ΔG, were calculated by Bardi et al.[22] We also found it very illuminating to have a close, comparative look at the HIV-1 mutant protease V82A,[34] in which Val82 is replaced by Ala. X-ray crystallographic data on its complex with the inhibitor A77003 is available (complex 1HVS). The mutant protease is inhibited by A77003 4-fold less than the native protease.

The C_2 symmetry content, $S(C_2)$, was calculated for each of the proteins amino acid pairs (AAPs), for each of the inhibitors before and after complexation, and for the whole protease in each complex. These calculations were carried out by applying the symmetry measure:

$$S(C_2) = \frac{100}{nD^2}\sum_{k=1}^{n}(p_k - \hat{p}_k)^2 \qquad (1)$$

where p_k is the vector column of the point coordinates of the analyzed shape (the input), $\hat{p}_k$ are the points coordinate of the nearest symmetric shape (the searched, computed output), n is the number of points, and D is the maximum distance from any point to the shape center. $S(C_2)$ measures the content of the point group $\{E, C_2\}$ in a structure; and the bounds are $0 \leq S \leq 100$, so that, as mentioned above, $S(C_2) = 0$ means perfect symmetry. The interested reader may wish to consult refs 3, 5−7, 20 for full details of the properties of this measure and for examples of its successful applications. Reference 18 also contains a discussion on the differences between the CSM methodology and rms methods.[35] Since X-ray data involves an uncertainty in the exact location of the atom, a situation which is particularly problematic for large objects such as proteins, we used here a method we had developed for computing the resulting confidence in the $S(G)$ values.[18] In general,

(11) For some recent HIV protease studies, published in this journal, see: (a) Hodge, C. N.; Lam, P. Y. S.; Eyermann, C. J.; Jadhav, P. K.; Ru, Y.; Fernandez, C. H.; De Lucca, G. V.; Chang, C.-H.; Kaltenbach, R. F., III; Holler, E. R.; Woerner, F.; Daneker, W. F.; Emmett, G.; Calabrese, J. C.; Aldrich, P. E. *J. Am. Chem. Soc.* **1998**, *120*, 4570−4581. (b) Lee, T.; Le, V.-D.; Lim, D.; Lin, Y.-C.; Morris, G. M.; Wong, A. L.; Olson, A. J.; Elder, J. H.; Wong, C.-H. *J. Am. Chem. Soc.* **1999**, *121*, 1145−1155. (c) Freedberg, D. I.; Wang, Y.-X.; Stahl, S. J.; Kaufman, J. D.; Wingfield, P. T.; Kiso, Y.; Torchia, D. A. *J. Am. Chem. Soc.* **1999**, *121*, 7916−7923.

(12) Babine, R. E.; Bender, S. L. *Chem. Rev.* **1997**, *97*, 1359−1472.

(13) Lapatto, R.; Blundell, A.; Hemmings, A.; Overington, J.; Wilderspin, A.; Wood, S.; Merson, J. R.; Whittle, P. J.; Danley, D. E.; Geoghegan, K. F.; Lee, S. E.; Scheld, K. G.; Hobart, P. M. *Nature* **1989**, *342*, 299−302.

(14) Greer, G.; Erickson, J. W.; Baldwin, J. J.; Varney, M. D. *J. Med. Chem.* **1994**, *37*, 1035−1054.

(15) Brickman, J. *Pure Appl. Chem.* **1997**, *69*, 55−69.

(16) Blundell, T. L.; Srinivasan, N. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 14243−14248.

(17) Bewley, C. A.; Gustafson, K. R.; Boyd, M. R.; Covell, D. G.; Bax, A.; Clore, G. M.; Gronenborn, A. M. *Nat. Struct. Biol.* **1998**, *5*, 571−578.

(18) Keinan, S.; Pinsky, M.; Plato, M.; Edelstein, J.; Avnir, D. *Chem. Phys. Lett.* **1998**, *298*, 43−50.

(19) Pinsky, M.; Avnir, D. *Inorg. Chem.* **1998**, *37*, 5575−5582.

(20) (a) Grimme, S. *Chem. Phys. Lett.* **1998**, *297*, 15−22. (b) Alikhanidi, S.; Kuz'min, V. *J. Mol. Model.* **1999**, *5*, 116−124.

(21) (a) Mezey, P. G.; Maruani, J. *Int. J. Quantum Chem.* **1993**, *45*, 177−187. (b) Cavalli, E.; Cammi, R. *Comput. Chem.* **1994**, *18*, 405−411. (c) Grant, J. A.; Gallardo, M. A.; Pickup, B. T. *J. Comput. Chem.* **1996**, *17*, 1653−1666.

(22) Bardi, J. S.; Luque, I.; Freire, E. *Biochemistry* **1997**, *36*, 6588−6596.

(23) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. R. J.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *J. Mol. Biol.* **1977**, *112*, 535−542.

(24) Sham, H. L.; Zhao, C.; Stewart, K. D.; Betebenner, D. A.; Lin, S.; Park, C. H.; Kong, X.-P.; Rosenbrook, W.; Herrin, T.; Madigan, D.; Vasavanonda, S.; Lyons, N.; Molla, A.; Saldivar, A.; Marsh, K. C.; McDonald, E.; Wdeburg, N. E.; Denissen, J. F.; Robins, T.; Kempf, D. J.; Plattner, J. J.; Norbeck, D. W. *J. Med. Chem.* **1996**, *39*, 392−397.

(25) Hosur, M. D.; Bhat, T. N.; Kempf, D. J.; Baldwin, E. T.; Liu, B.; Gulnik, S.; Wideburg, N. E.; Norbeck, D. W.; Appelt, K.; Erickson, J. W. *J. Am. Chem. Soc.* **1994**, *116*, 847−855.

(26) Kim, E. E.; Baker, C. T.; Dwyer, M. D.; Murcko, M. A.; Rao, B. G.; Tung, R. D.; Navia, M. A. *J. Am. Chem. Soc.* **1995**, *117*, 1181−1182.

(27) Thompson, S. K.; Murthy, K. H. M.; Zhao, B.; Winborne, E.; Green, D. W.; Fisher, S. M.; DesJarlais, R. L.; Tomaszek, T. A.; Meek, T. D.; Gleason, J. G.; Abdel-Meguid, S. S. *J. Med. Chem.* **1994**, *37*, 3100−3107.

(28) Priestle, J. P.; Fassler, A.; Rosel, J.; Intelnot-Blomley, M. *Structure (London)* **1995**, *3*, 381−389.

(29) Abdel-Meguid, S. S.; Metcalf, B. W.; Carr, T. J.; Demarsh, P.; DesJarlais, R. L.; Fisher, S.; Green, D. W.; Ivanoff, L.; Lambert, D. M.; Murthy, K. H. M.; Petteway, S. R.; Pitts, W. J.; Tomaszek, T. A.; Winborne, E.; Zhao, B.; Dreyer, G. B.; Meek, T. D. *Biochemistry* **1994**, *33*, 11671−11677.

(30) Thaisrivongs, S.; Watenpaugh, K. D.; Howe, W. J.; Tomich, P. K.; Dolak, L. A.; Chong, K.-T.; Tomich, C.-S. C.; Tomasselli, A. G.; Turner, S. R.; Strohbach, J. W.; Mulichak, A. M.; Janakiraman, M. N.; Moon, J. B.; Lynn, J. C.; Horng, M.-M.; Hinshaw, R. R.; Curry, K. A.; Rothrock, D. J. *J. Med. Chem.* **1995**, *38*, 3624−3637.

(31) Hong, L.; Treharne, A.; Hartsuck, J. A.; Foundling, S.; Tang, J. *Biochemistry* **1996**, *35*, 10627−10633.

(32) Hoog, S. S.; Zhao, B.; Winborne, E.; Fisher, S.; Green, D. W.; DesJarlais, R. L.; Newlander, K. A.; Callahan, J. F.; Moore, M. L.; Huffman, W. F.; Abdel-Meguid, S. S. *J. Med. Chem.* **1995**, *38*, 3246−3252.

(33) Since some of the data includes hydrogens while other do not, we removed all hydrogens when available, for conformity of all analyses.

(34) Baldwin, E. T.; Bhat, T. N.; Liu, B.; Pattabiraman, N.; Erickson, J. W. *Nat. Struct. Biol.* **1995**, *2*, 244−249.

(35) Kabsch, W. *Acta Crystallogr.* **1976**, *A32*, 922−923.

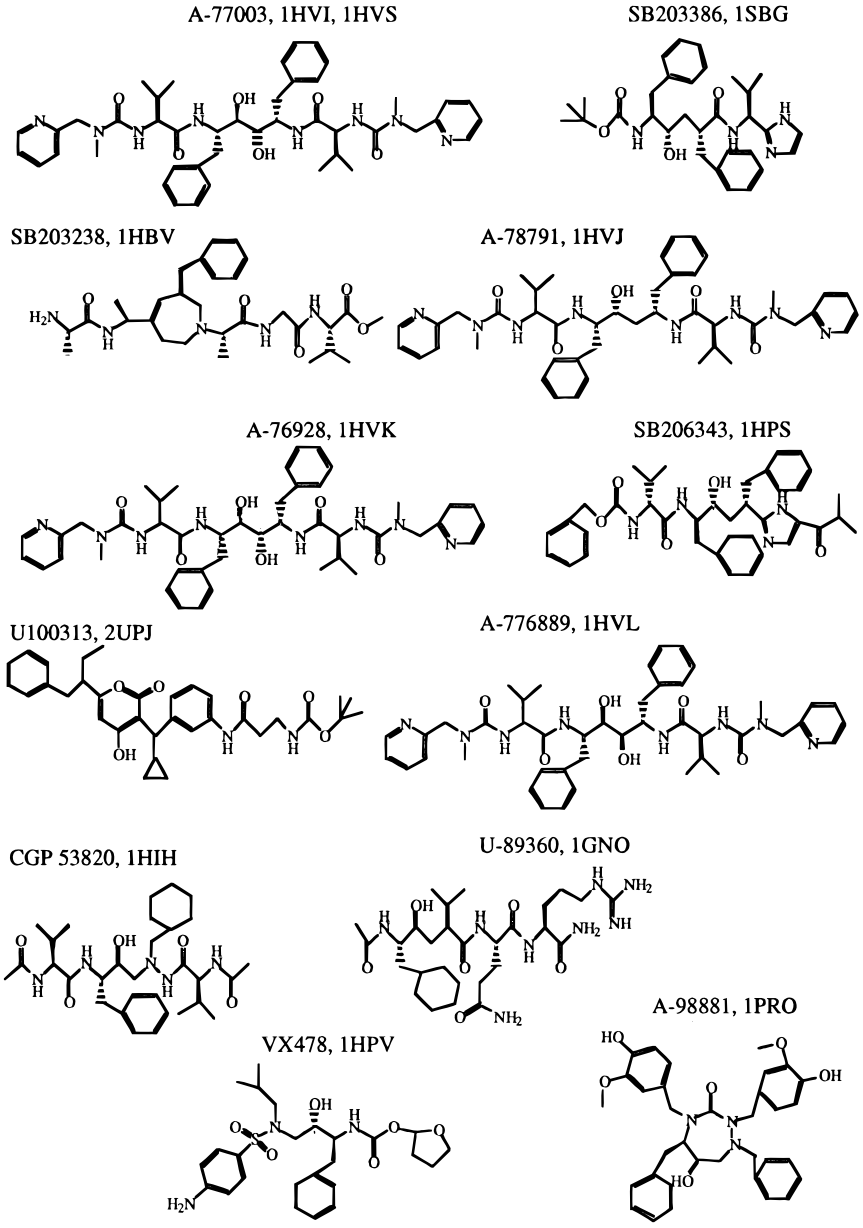

Figure 1. The twelve inhibitors used in this study. For each structure, the left code is of the inhibitor and the right one is the PDB code for the protease−inhibitor complex.

for the large proteins, the error bar in $S(C_2)$ is of the order of ± 0.0005, while for the small inhibitors (and for the AAPs of the active site[36]) it is of the order of ± 0.002. (Since many of the residues, which we identify below as highly symmetric-sensitive, lie on the surface or in the flap region, the $S(C_2)$ values for these residues might reflect the crystallographic *B* factors of these flexible side chains. However, no correlation was found between the specific $S(C_2)$ values of AAPs and their crystallographic *B* factors (data available from the authors), and so this seems to corroborate the authenticity of $S(C_2)$ values as representing inherent deviations of the AAPs).

The free, unbound minimum energy structures were determined by semiempirical AM1 geometry optimizations, as embedded in the Spartan program,[37] employing the gradual descent as an energy minimization algorithm. The input structures were the bound, complexed structures from the X-ray data. The rms values of the 1HVI complex (average of the two amino acids within each AAP) as compared to the native protein (3PHV), were calculated by ProFit V1.8.[38]

Results and Discussion

Inhibitor and Active-Site Symmetry Changes. Table 1 lists the $S(C_2)$ values of the inhibitors before and after complexation. Figure 2 reveals a correlation between these two structural features: For most inhibitors the complexation forces the structure to be *nearer* to C_2 symmetry, compared to the free form. Increase in deviation upon binding, whenever it occurs, is always small, with $\Delta S(C_2)$ not exceeding 0.1. Among those inhibitors which experience deviation increase are the two fully C_2-symmetric inhibitors (the $S(C_2) = 0.0$ entries in Table 1), and since the symmetry of the active site is also perfectly C_2-symmetric prior to binding, the positive $\Delta S(C_2)$ is indicative of the asymmetric docking of these inhibitors within the active site.[25]

(36) Erickson, J.; Neidhart, D. J.; VanDrie, J.; Kempe, D. J.; Wang, X. C.; Norbeck, D. W.; Plattner, J. J.; Rittenhouse, J. W.; Turon, M.; Wideburg, N.; Kohlbrenner, W. E.; Simmer, R.; Helfrich, R.; Paul, D. A.; Knigge, M. *Science* **1990**, *249*, 248−533.

(37) *Spartan*, DEC version 4.1.2 X11 ed.; Wavefunction Inc.: Irvine, CA, 1995−1999.

(38) Martin, D. A. C. R. *ProFit*; 1.8 ed.; SciTech Software, 1992−1998.

Table 1. The C_2 Symmetry Properties of HIV Protease−Inhibitor Complexes, with their ΔG Values

PBD entry	inhibitor (Figure 1)	$S(C_2)$ free inhibitor (calculated)	$S(C_2)$ complexed inhibitor	$\Delta[S(C_2)]$ for the inhibitor	$S(C_2)$ active site[a]	$S(C_2)$ protein−inhibitor complex	ΔG binding energy [cal/mol]
1PRO	A98881	11.346	11.438	0.093	0.097	0.0246	−15417
1HVJ	A78791	0.005	0.103	0.098	0.104	0.0200	−14299
1HVK	A76928	0.000	0.017	0.017	0.082	0.0180	−13796
1HVI	A77003	0.308	0.140	−0.168	0.118	0.0190	−13699
1HVS[b]	A77003	0.308	0.125	−0.183	0.099	0.0200	−12914
1HPV	VX478	8.288	8.388	0.100	0.137	0.0285	−12580
1HPS[c]	SB206343	6.353	6.154	−0.199	0.306	0.0275	−12580
1HVL	A76889	0.000	0.023	0.023	0.085	0.0261	−12278
1HIH	CGP53820	13.277	12.201	−1.076	0.147	0.0312	−10976
1SBG	SB203386	12.343	11.297	−1.046	0.234	0.0317	−10565
2UPJ	U100313	6.341	4.466	−1.875	0.201	0.0214	−10077
1GNO[c]	U89360	9.902	7.678	−2.224	0.129	0.0235	−9441
1HBV	SB203238	10.927	6.447	−4.480	0.102	0.0279	−8685

[a] The active site, as defined in ref 36, constitutes of the AAPs in S1: Leu23, Asp25, Gly27, Ala28, Gly49, Ile50, Val82, Ile84. [b] The mutant. [c] For two complexes 1GNO and 1HPS, there are two possible positions of the inhibitor in the active site. We use the average.

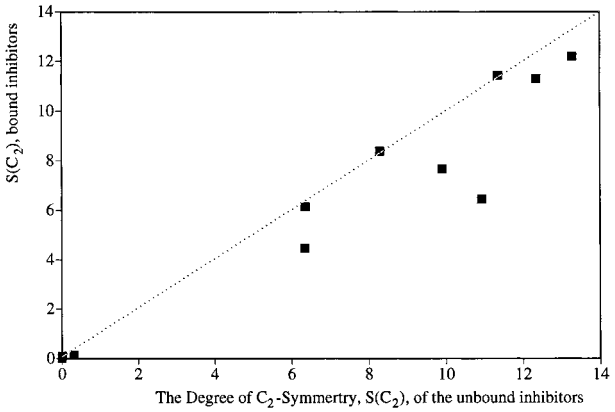

Figure 2. The C_2-symmetry content values ($S(C_2)$) for the bound, complexed inhibitors (as obtained from X-ray data) vs the $S(C_2)$ values of the structures (calculated) of the unbound inhibitors. The dashed line represents the hypothetical case of no symmetry change upon binding: Note that the active site either imposes better C_2-symmetry on the bound inhibitor or changes it very little.

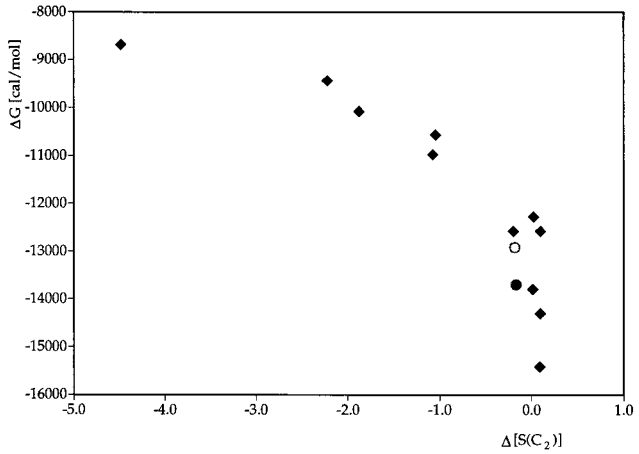

Figure 3. Free energy of inhibitor binding (ΔG[16] vs the symmetry change impose of the inhibitor upon binding ($\Delta S(C_2)$). Circles: the two complexes with inhibitor A77003: (●) the complex with the native protease (1HVI); (○) with the V82A mutant protease (1HVS).

Having in mind the nonrigidity of the active site, the asymmetric docking must also affect, in turn, the symmetry of the active site as well. It is seen in Table 1 that in fact, all of the inhibitors induce a distortion in the active-site original C_2 symmetry upon complexation. In all cases the $\Delta S(C_2)$ value of the active site is smaller than that of the inhibitor, reflecting the smaller conformational adjustability of the active site (and of the whole protein, see below) compared to the inhibitor.

Symmetry-Change Correlation with Binding Affinities. Since structural changes of the inhibitors upon binding are accompanied by energy changes, it was interesting to see if the two—a shape-change parameter and a thermodynamic-change parameter—are linked in anyway. Figure 3 shows, we believe for the first time, a clear correlation between symmetry changes upon binding, $\Delta S(C_2)$ in this case, and the experimental binding affinities, ΔG[22] (Table 1). A trend is evident: Higher binding affinities (higher negative values) are associated with lower $\Delta S(C_2)$ values. Through $\Delta G = -RT \ln K_i$, this also means more efficient inhibition, as $\Delta S(C_2)$ decreases. There is a delicate, very interesting point to notice here: It is not that the better symmetry fit is associated with better binding; rather, the message of the correlation we revealed is that *strong binding requires that the change in symmetry of the inhibitor be minimal*. In other words, high ΔG values of interaction are gained whenever the enzymes does not invest much in distorting the minimal-energy free conformation of the inhibitor.

Symmetry Profiles and Symmetry Maps. As much as the whole protein takes part in the engineering of the geometry of the active site, and thus variations in it will back-affect the protein. Thus, it is expected that the deviation from C_2 symmetry which originates from the perturbation of the active site, will spread throughout the whole protein. In what way? What is the pattern of the distribution of the C_2 deviation within the protein?

To answer these important questions, we introduce here two new structural/analytical observation tools. The first is the *symmetry sensitivity profile*, in which the S(C_2) value of each AAP is plotted against its position in the backbone, as shown in Figure 4a (for the 1HVI complex). To understand the unique feature provided by the symmetry profile, let us compare it to a common representation of distortion, namely an rms analysis. This is shown in Figure 4b in which the rms distance between the original location of an AAP in the protease before complexation (the C_2-symmetric 3PHV[13]) and its location in the complex is plotted against its backbone position. It is seen that *all* AAPs move, however, no information is provided on the *symmetry* distortion due to the motion: At an extreme, within a given AAP, the two AAs can move identically, resulting in zero $\Delta S(C_2)$. Thus, Figure 4a distills out the asymmetry involved in the distortive motion. In comparing Figure 4a with 4b it is seen, however, that quite often, large motion of an AAP (high rms) is carried out in a nonsymmetric way (large $S(C_2)$ value).

Having in mind the information provided by the symmetry sensitivity profile, we compare now the profile of 1HVI (Figure

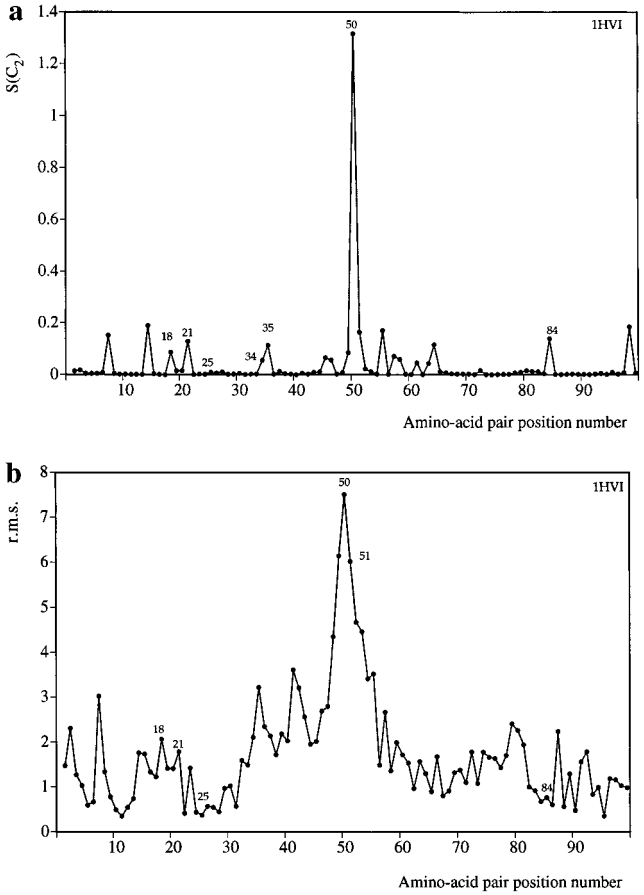

Figure 4. (a) The symmetry sensitivity profile of the protease, complexed to the inhibitor A77003 (1HVI). The $S(C_2)$ value for each of the AAPs is plotted vs its position. (Numbered are AAPs mentioned in the text). For comparison (see text), the rms values are given in (b).

4a) to the profiles (Figure 5) of three other complexes: 1HVK which is a complex with a perfectly C_2-symmetric inhibitor in its free form (A-76928), Figure 5a; 1SBG, Figure 5b; and 1HBV, Figure 5c. The first two, 1HVI and 1HVK have low $\Delta S(C_2)$ values, negative and positive, respectively; 1SBG is in the middle of the $\Delta S(C_2)$ range; and 1HBV has the highest $\Delta S(C_2)$ value in our series of complexes. It is seen that the symmetry sensitivity profiles have two functions: They can serve as comparative fingerprints, each one unique for a specific protease−inhibitor complex; and they can help in identifying common features among the various inhibitor complexes. Thus, for all four complexes, a feature which immediately strikes the eye, is the high symmetry deviation of the Ile50 residues which are at the "flap region" (shown in Figure 6.). In fact, the HIV protease shows high symmetry sensitivity of this AAP for *all* of the inhibitors of Figure 1. Sub-families have their unique features as well: for instance, 1HVI (Figure 4a), 1HVK (Figure 5a), 1HVS, 1HVJ, and 1HVL, a family of protease complexes with similar diol inhibitors, show a unique double peak at Gln18-Glu21, a combined peak of Glu34-Glu35, and a peak at Ile84. We return to these profiles in the analysis of an HIV-1 mutant, below.

Since long-range structural effects in proteins do not follow the backbone but are operative through space, it is also of interest to determine how are the symmetry deviations distributed *spatially* within the protein. For that purpose we introduce the second structural observation tool, namely the *symmetry sensitivity map*, shown in Figure 6 for the complex 1HVI. The map depicts at one glance the symmetry deviation pattern of

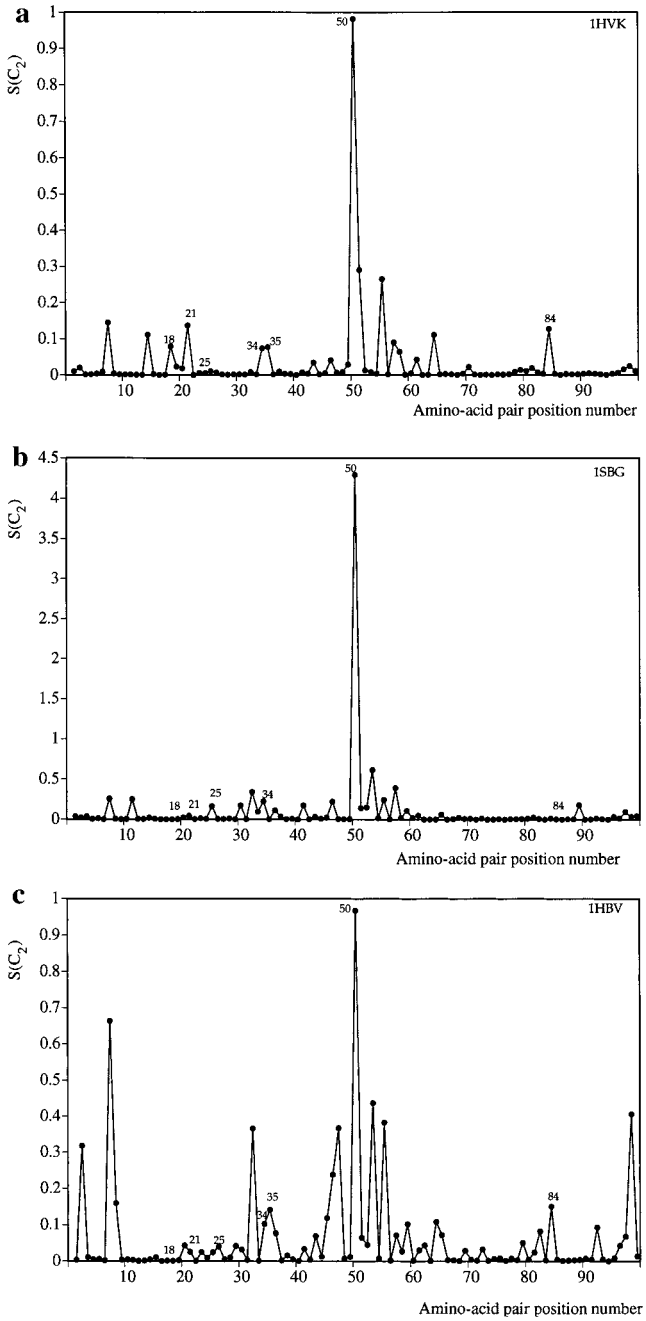

Figure 5. The symmetry sensitivity profiles of (a) 1HVK (inhibitor A76928), (b) 1SBG (inhibitor SB203386), (c) 1HBV (inhibitor SB203238).

the protein by using various colors for various $S(C_2)$ levels. It is immediately evident that large symmetry deviations (red and green) are not associated necessarily with the active site, but are scattered at various locations of the protein complex. It is also seen which of the protein's contact points with the inhibitor (A77003) are symmetry-affected, and which are not: One contact point is with one of the flap's AAPs, Ile50 (through a hydrogen bond between the nitrogen on the Ile50 and a bridging water molecule to a carboxylic oxygen on the inhibitor), and that AAP, as we have already seen, deviates significantly. Another contact point is with one of the Asp25 AAPs of the catalytic triad (again through a hydrogen bond), but this one is barely affected, and so are the rest of the AAPs at the vicinity of the inhibitor. Interestingly, this small symmetry deviation of the AAPs in the active-site region was found to be the case even for the most distorted complex, 1SBG (Figure 7). Thus,

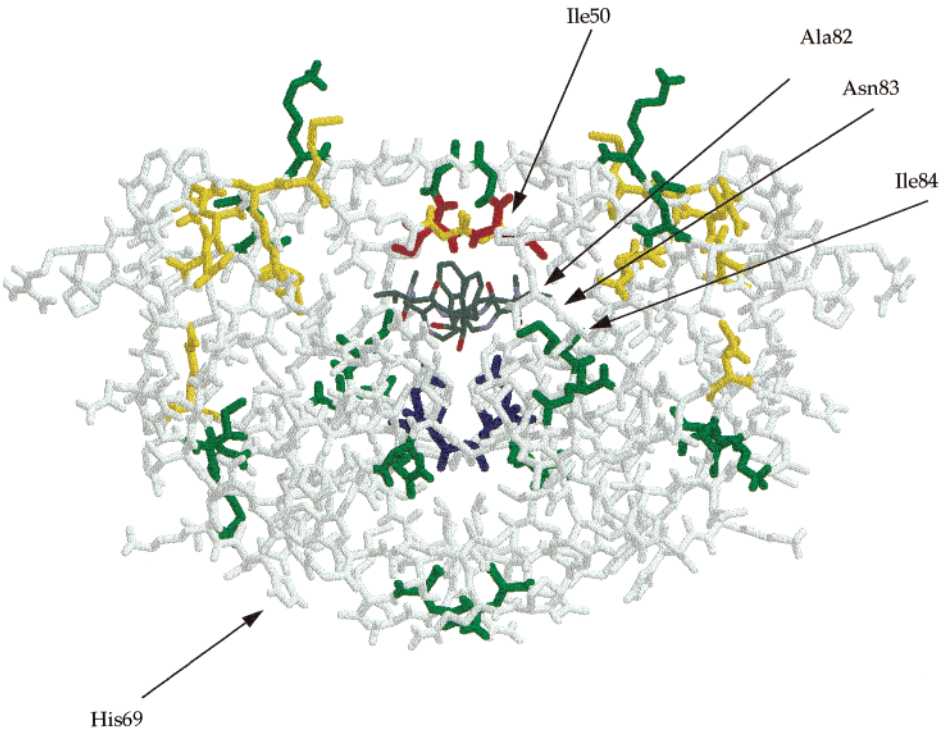

Figure 6. The symmetry sensitivity map of 1HVI. Red represents the highest symmetry deviation (Ile50 $S(C_2) = 1.31$), green represents symmetry deviations in the range $1 > S(C_2) > 0.1$, yellow are deviations in the $0.1 > S(C_2) > 0.05$ range, and light gray represents very small deviations (below 0.05). The catalytic triad (Asp25, Thr26, Gly27) is shown in cyan, and the inhibitor is seen above it in dark gray. In the mutant 1HVS, AAP 82 is replaced and the symmetry of AAPs 69,82,83 is consequently affected.

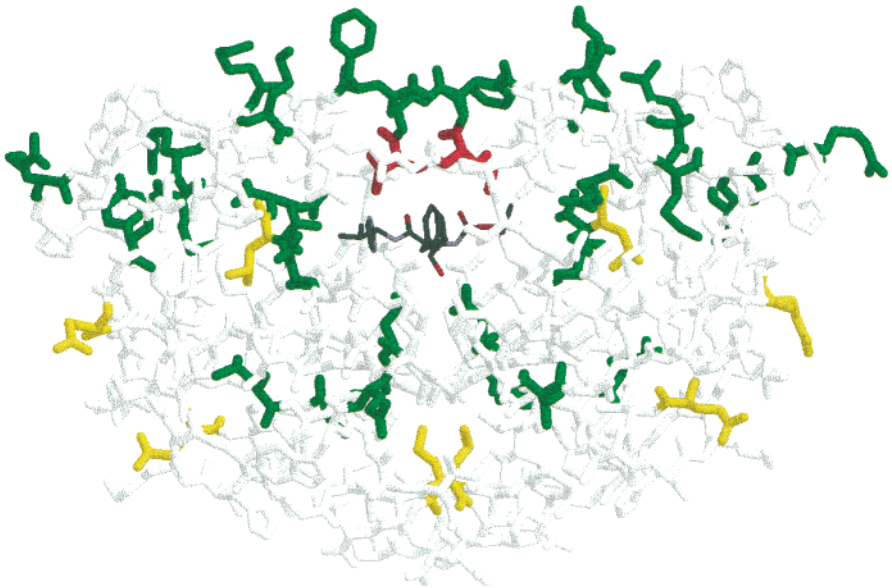

Figure 7. The symmetry sensitivity map of 1SBG. Red: Ile50 $S(C_2) = 4.29$; green: $1 > S(C_2) > 0.1$; yellow: $0.1 > S(C_2) > 0.05$ and light gray represents $S(C_2) < 0.05$. The inhibitor is in dark gray.

an important property of enzymes response is evident here: Small symmetry variations for most of the AAPs that surround the inhibitor, are translated into large symmetry changes in remote regions of the enzyme; it is a manifestation of the *action of the enzyme as a mechanical ("scissors") device.*

The Mutant Protease V82A. Fast mutations of HIV in general, and of the protease structure in particular, are one of the major obstacles toward achieving efficient therapy of AIDS. Specifically, we focus on the mutant strain V82A which was raised against inhibitor A77003 (and against other inhibitors).[34] The mutated protease of this virus has an Ala instead of a Val at the 82 AAP, which is part of the active site. Here we show how the symmetry analysis follows and interprets in a novel way the adaptation of a mutated protease to an inhibitor. For this we analyze the complex between the mutated protease and inhibitor A77003 (1HVS) and compare it to the nonmutated analogue described above, 1HVI.

As seen in the symmetry profile of 1HVS (Figure 8a, compared with that of 1HVI Figure 4a), the mutation, which is of only *one* AAP, casts changes in the way the symmetry distortion spreads out in the protein. This is seen in an even clearer way in the symmetry difference profile, Figure 8b. Interestingly, although the AAP82 mutation barely affects the symmetry of that site (in 1HVI the inhibitor interacts with

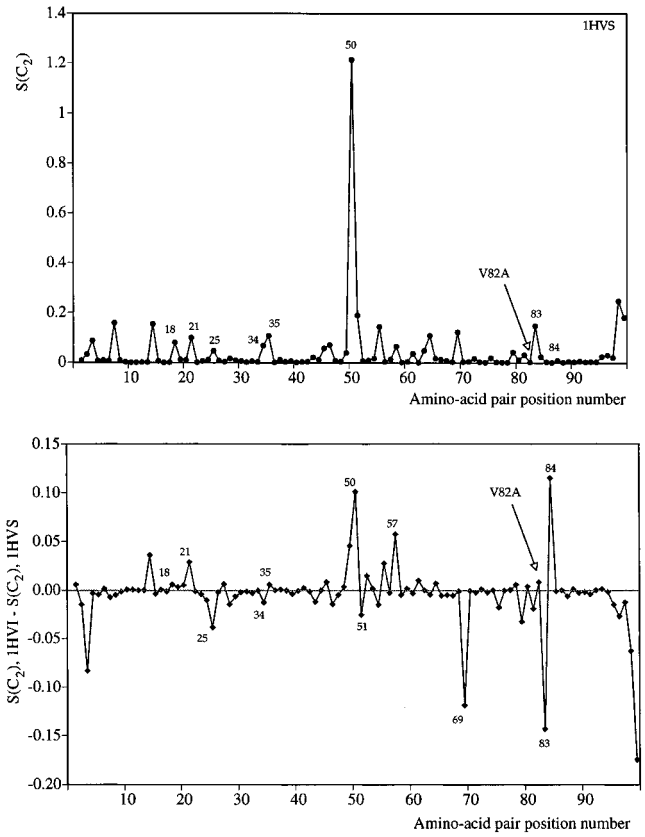

Figure 8. (a) The symmetry sensitivity profile of the mutant protease complex 1HVS (inhibitor A77003) and (b) the effect of the mutation on the symmetry: the difference between the profiles of 1HVI and 1HVS.

AAP82 while in the mutant it does not) its immediate neighbors, Asn83 and Ile84, are affected markedly (increase of the former and decrease of the latter; note also the marked effect on His69—see the location of all of these AAPs in 1HVI Figure 6).

As for the response of the mutant to the inhibitor, if the correlation we found in Figure 3 is predictive, then this means, first, that the mutant should "sit" on the same curve of ΔG vs $\Delta S(C_2)$; and second, that it should sit there *higher* than the native one. Remarkably, this is indeed the case (Figure 3): The mutation increases the $\Delta S(C_2)$ value from -0.168 to -0.183, accompanied by a decrease in the ΔG absolute value from 13699 cal/mol to 12914 cal/mol. As seen in Figure 3, the HIV has "learned" and has made—in a figurative expression—"the right decision": By inducing the V82A replacement, it has increased the absolute value of $\Delta S(C_2)$, which, as we have seen above, results in moving up the curve of Figure 3, namely in being less affected by the inhibitor, compared to the native protease. We regard this observation as one of the key results of this study.

The Symmetry Deviation of the Whole Protease. Since the whole of the protein takes part in sculpturing the structure and properties of the active site, we take now a closer look at the symmetry changes the protease undergoes upon inhibition. Table 1 summarizes the overall protein deviations caused by the complexation. It is remarkable that despite wide variations in the $S(C_2)$ and $\Delta S(C_2)$ values of the inhibitors, *the level of C_2 distortion of the protein as a whole is practically unaffected by the nature of the inhibitor,* and stands on an average value of 0.025 ± 0.005. Our tentative proposition for this intriguing observation is that one is witnessing here the known filtering behavior of large dynamic systems.[39] Such systems are capable of translating various types of input signals into a stable, system-characteristic output. This behavior is in particular evident in complex systems with plenty of internal correlations and feedback loop responses. Thus, the proposition is that the protease has more than one symmetry minimum (the perfect symmetry of the apo protein, except for one water molecule at the active site): An additional "symmetry-well" shows up in the 0.025 value, once an inhibitor enters the active site. To test this idea, we took 14 other protease-inhibitor complexes randomly selected from the PDB, and we found the two symmetry minima again: Four complexes gave the perfect symmetry value (1HEF, 1HEG, 1HHP, $S(C_2) = 0.000$ and 1AAQ, $S(C_2) = 0.001$), and all other 10 complexes gave an average of $S(C_2) = 0.024(!) \pm 0.005$ (A8G, $S(C_2) = 0.018$; 1AID, $S(C_2) = 0.031$; 1AJV, $S(C_2) = 0.030$; 1AJX, $S(C_2) = 0.026$; 1DIF, $S(C_2) = 0.021$; 1HII, $S(C_2) = 0.019$; 1HIV, $S(C_2) = 0.029$; 1HOS, $S(C_2) = 0.019$; 1HPX, $S(C_2) = 0.021$; 1HSG, $S(C_2) = 0.020$).

We are not ready yet to comment on the possible relation between this global symmetry minimum and energy minimum. One should take into account that, like energy, symmetry is a global descriptor which is indifferent, within bounds, to specific geometry details which characterize each complex. Yet hints on such symmetry/energy relation do exist: Recently we found, both for isolated molecules[9,10] and for large assemblies of Lenard-Jones clusters[40] that symmetry and energy correlate nearly linearly. We regard the identification of the constant symmetry deviation of protease complexes as another key result of this study, the full consequences of which are still awaiting to be unveiled.

Concluding Remarks

We have demonstrated that symmetry analysis can be a powerful tool in the structural analysis of proteins, of their complexes, and of their performances. Symmetry sensitivity profiles and symmetry maps provide at a glance the full information for the whole protein. Symmetry analysis provided here the first correlation between this structural property and binding, demonstrating the importance of global shape descriptors for active site interactions, in distinction from specific structural parameters. This corroborates a similar conclusion we reached from the quantitative chirality analysis of inhibition with chiral substrates.[1] The predictive value of the symmetry/energy correlation was demonstrated for the mutant behavior, which "learned" to follow this correlation. We thus believe that the symmetry (and chirality) performance correlations could prove to be a useful novel paradigm in the design of inhibitors and drugs.

Last but not least, the complex machine-type of behavior of the enzyme showed up not only in identifying a common symmetry minimum of HIV protease to which it resorts upon binding to an inhibitor, but also in the way the symmetry distortion upon inhibition spreads out from the active site into the protein; and by the fact that for most inhibitors, the active site imposes a nearer C_2-symmetry on the inhibitor than its symmetry in the free form.

Acknowledgment. This work was supported by the U.S.—Israel Binational Science Foundation and by a Horowitz Fund. S.K. thanks the J. Klein Foundation for an award.

JA993810N

(39) Bezrukov, S. *Phys. Lett. A* **1998**, *248*, 29—36.
(40) Pinto, Y. Ph.D.; The Hebrew University of Jerusalem, Jerusalem, 1999.