

Utility of the Hard/Soft Acid–Base Principle via the Fukui Function in Biological Systems

John Faver and Kenneth M. Merz, Jr.*

Quantum Theory Project, Department of Chemistry, University of Florida, 2328 New Physics Building, P.O. Box 118435, Gainesville, Florida 32611-8435

Received September 25, 2009

Abstract: The hard/soft acid–base (HSAB) principle has long been known to be an excellent predictor of chemical reactivity. The Fukui function, a reactivity descriptor from conceptual density functional theory, has been shown to be related to the local softness of a system. The usefulness of the Fukui function is explored and demonstrated herein for three common biological problems: ligand docking, active site detection, and protein folding. In each type of study, a scoring function is developed on the basis of the local HSAB principle using atomic Fukui indices. Even with necessary approximations for its use in large systems, the Fukui function remains a useful descriptor for predicting chemical reactivity and understanding chemical systems.

1. Introduction

Computational biochemistry is an expanding field that relies heavily on the increasing efficiency of computers and clever algorithms to approach very large and complex problems. While methods for high-level quantum mechanical (QM) calculations have been developed and proven to be very successful in calculating energies, equilibrium structures, vibrational frequencies, and more properties of small- to medium-sized molecules, the computational resources required for very large systems (e.g., a protein of many hundreds of atoms) are usually unattainable.^{1–3} For these very large systems, more approximate modeling tools are often used such as molecular mechanics.^{4,5} These more approximate methods greatly accelerate the speed at which energy calculations are performed, but they do not explicitly account for electronic structure. This can be a disadvantage, because there is a significant amount of information encoded in the electronic structure of a system.

Conceptual density functional theory (CDFT) defines many reactivity descriptors for a system based on its electron density and provides a large set of tools for use in the prediction and understanding of chemical reactivity. An extensive review of CDFT and the myriad of possible descriptors has been compiled by Geerlings, De Proft, and Langenaeker.⁶ These descriptors have been used in the past

for a diverse set of chemical systems.^{7–9} More recently, they have been used with some success in biochemically relevant systems including the detection of metabolic sites in known drug molecules, the understanding of metal binding to porphyrin, and enzymatic catalysis.^{10–12} A beneficial characteristic of these descriptors is that the majority of them depend on quantities such as electron density that can be obtained from any QM method, including semiempirical QM Hamiltonians.^{13,14}

In the past two decades, advances in algorithms have allowed computational chemists to perform QM calculations on large systems such as proteins.^{15,16} One such method is the divide and conquer method.^{17–21} By dividing a molecule into smaller subsystems and performing separate calculations followed by the formation of a global density matrix, the method greatly accelerates calculations for large systems. An important result of this development is that electron density and descriptors based on electron density can now be calculated for large molecules as well as small molecules. Khandogin and York recently described a few such useful descriptors for divide and conquer semiempirical calculations.²²

Pearson's hard/soft acid–base (HSAB) principle states that chemical species can be described as being either hard or soft acids or bases.²³ Soft species tend to be easily polarizable and large in volume, have low charge, and have small HOMO–LUMO gaps. Hard species tend to have the opposite characteristics: they are not easily polarized, are small in

* Corresponding author phone: (352) 392-6973; fax: (352) 392-8722; e-mail: merz@qtp.ufl.edu.

volume, are highly charged, and have large HOMO–LUMO gaps. The HSAB concept can be summarized as one simple rule: hard species favor interacting with hard species, and soft species tend to favor interacting with soft species. The HSAB concept has been successful in predicting reactivity preferences in many systems since its inception.^{24–32}

Researchers have devised various methods of quantifying hardness and softness. Although empirical approximations have been used in the past, in this paper we will describe the use of one related reactivity descriptor from CDFT called the Fukui function which has been shown to carry information about chemical softness.^{33–35} This work then explores its applicability to biological problems, specifically ligand docking, active site detection, and protein folding.

2. Background

According to density functional theory, changes in the electronic energy $dE[\rho(r)]$ are related to changes in the number of electrons N and changes in the external potential $v(r)$ felt by the electron distribution (which usually refers to the nuclear positions in chemical systems):

$$dE[\rho(r)] = \mu dN + \int \rho(r) dv(r) dr \quad (1)$$

For simplicity, consider a molecule at a given geometry in its ground state so that $dv(r)$ is zero. Thus, the partial derivative of energy with respect to the number of electrons N at constant geometry is the electronic chemical potential μ :

$$\mu = \left(\frac{\partial E}{\partial N} \right)_{v(r)} = -\chi \quad (2)$$

This quantity has been related conceptually to the electronegativity χ of a system.³⁶ This definition agrees with chemical intuition, as more energetically favorable changes in electron number yield higher values of electronegativity. Consider now the second partial derivative of the energy with respect to the electron number

$$\eta = \left(\frac{\partial^2 E}{\partial N^2} \right)_{v(r)} = \left(\frac{\partial \mu}{\partial N} \right)_{v(r)} \quad (3)$$

which has been defined as η , or chemical hardness as described by Pearson.³³ This definition can be understood by the analogy of a spring constant in classical physics. The spring constant is the second derivative of energy with respect to displacement and measures the difficulty of displacing a spring from its equilibrium position. Equation 3 can be thought of as measuring the difficulty of changing a system's number of electrons, which is conceptually similar to nonpolarizability, or hardness. Since softness is the opposite of hardness, it has been defined as the inverse of hardness:

$$S = \frac{1}{\eta} = \left(\frac{\partial N}{\partial \mu} \right)_{v(r)} \quad (4)$$

Parr and Yang have also defined a distance-dependent version of softness, called the local softness, as

$$s(r) = \left(\frac{\partial \rho(r)}{\partial \mu} \right)_{v(r)} = \left(\frac{\partial \rho(r)}{\partial N} \right)_{v(r)} \left(\frac{\partial N}{\partial \mu} \right)_{v(r)} = f(r) S \quad (5)$$

The local softness function identifies the softest regions of a molecule. A system has a total softness S that is distributed throughout the molecule by a function $f(r)$ called the Fukui function:

$$f(r) = \left(\frac{\partial \rho(r)}{\partial N} \right)_{v(r)} = \frac{s(r)}{S} \quad (6)$$

The Fukui function is normalized to unity so that the local softness integrates over all space to yield the total softness. Furthermore, the Fukui function can be viewed as containing the same information as the local softness, since the two are proportional to each other by a constant, S . Although there exist several descriptors for local hardness, the problem of defining it has not been resolved.^{37,38} In this work, low values of local softness are assumed to be locally hard. From the equations of DFT, we now have the Fukui function, a descriptor that identifies the softest (and hardest) regions of a molecule. With this knowledge in hand, one can begin to make predictions about chemical reactivity.

One issue that arises when calculating the Fukui function is that it is a derivative of the electron number, which is by nature an integer. Although recent studies have examined ways to circumvent this apparent discontinuity, these methods are impractical at this time for the large systems considered here.^{39,40} Limiting the calculations to changes with integer electrons, it is necessary to use finite difference derivatives. With the finite difference formulas, there is the option of taking the derivative from the left, right, or center:

$$f(r)^- \cong \frac{1}{\Delta N} [\rho(r, N) - \rho(r, N - \Delta N)] \quad (7)$$

$$f(r)^+ \cong \frac{1}{\Delta N} [\rho(r, N + \Delta N) - \rho(r, N)] \quad (8)$$

$$f(r)^0 \cong \frac{1}{2} [f(r)^+ + f(r)^-] \quad (9)$$

The Fukui function taken from the left is the difference in electron density between the reference system and the system with an electron removed, e.g., a ground state and its cation (eq 7). Because maxima in this function represent areas where electron density is most favorably decreased, they are interpreted as areas in a molecule most favorable for electrophilic attack. The Fukui function taken from the right has maxima that are interpreted as areas most favorable for nucleophilic attack, since it detects areas where electron density increases most favorably under addition of electrons (eq 8). The centered derivative is simply the average of the two other derivatives and has often been interpreted as showing areas most favorable for attack by a radical (eq 9). A recent study has explored the validity of this interpretation and found that it may not be quite as easy to interpret as the other derivatives.⁴¹ While the left and right derivatives are clearly understood in terms of two classical reaction mechanisms, the middle derivative can for now be viewed as the best approximation of the derivative at the reference state.

In addition to finite difference derivatives, a second common approximation of the Fukui function is the condensed Fukui function, which is composed of atomic Fukui indices.⁴² Within this approximation, atomic partial charges are used to replace the electron density in the expression for the Fukui function. Though this may be a crude approximation of the full electron density and Fukui function, several studies have been successful with its use.^{7,8,10,12,43–45} In general, one must choose a density partitioning scheme which unfortunately can depend heavily on the QM method or basis set and thus introduce error. Because of this, Fukui indices are sometimes negative, which seems unphysical. A negative Fukui index implies that addition of electrons to a system decreases density in locations in the system or vice versa. Though some example molecules have been shown to have this interesting property, it should not be as common as the use of Fukui indices suggests.^{46,47} Keeping this in mind, Fukui indices were used in this work rather than full Fukui functions, simply because full Fukui functions are considerably more expensive to calculate.

A third common approximation is the use of the frozen orbital approximation, in which a single calculation is done to obtain the eigenstates of the system which are assumed to be “frozen” in place as electrons are added or removed. Clearly, changing the electron number in a system will alter the forces felt by the remainder of the electrons, and the eigenstates will be altered in a phenomenon called orbital relaxation. Examples have been shown in which Fukui indices based on the frozen orbital approximation fail to predict correct reactivity in small organic molecules.⁴⁷ Orbital relaxation effects were taken into account in this work by performing separate calculations for the ground-state system, the system with added electrons, and the system with electrons removed, rather than using the frozen orbital approximation.

Khandogin et al. have described the calculation and interpretation of several QM-based reactivity descriptors for biomolecules, including the Fukui function and local softness.²² The present work concentrates on one of them, the Fukui function, and attempts to determine in what kinds of applications it can be used and for what kinds of interactions it can account and then determine the extent of its reliability. With these approximations (divide and conquer, AM1, finite difference, Mulliken atomic charges), five specific types of problems were addressed: finding correct ligand poses in an active site, detecting active binders from a set including decoy ligands, ranking binding affinities of ligands, finding reactive sites in a protein, and detecting native from decoy protein structures. In each of these systems, it was hypothesized that molecular interactions are favorable when hard areas are near hard areas and soft areas are near soft areas. In each of the five problems, a scoring function based on this hypothesis is developed and used to predict the preferred molecular interactions.

3. General Computational Details

All calculations were performed with the semiempirical AM1 Hamiltonian in the DivCon program utilizing the divide and conquer strategy for proteins.^{17–19,48} Standard unrestricted

AM1 calculations were done in DivCon for all ligand molecules. Unrestricted divide and conquer calculations were done for the proteins in the 1F40 and 2FOM docking studies, and restricted divide and conquer calculations were done for the 1EFY docking study and the 1ORC and 1I6C protein folding study. Atomic Fukui indices were calculated from centered finite difference derivatives of Mulliken charges. The derivatives were calculated by varying the electron number by 1 for ligand molecules. The electron number was varied by 5 in the case of 1F40 and 2FOM, by 8 for 1EFY, and by 4 for 1ORC and 1I6C. These values were chosen to roughly correspond to the size of the proteins. Fukui indices and molecular surfaces were visualized with the program PYMOL.⁴⁹

3.1. Docking. Docking studies were performed on three protein/ligand systems: FKBP12 (PDB ID 1F40), dengue virus type 2 NS3 protease (PDB ID 2FOM), and poly(ADP-ribose) polymerase (PDB ID 1EFY). 1F40 is an NMR structure bound to a synthetic ligand, GPI-1046.⁵⁰ 2FOM is an X-ray crystal structure (1.50 Å resolution) without a bound ligand.⁵¹ Several known active binders with experimental IC₅₀ values for the dengue protease were taken from a previous docking study.^{52,53} 1EFY is an X-ray crystal structure (2.20 Å resolution) with bound inhibitor. The structure was taken from the DUD (directory of useful decoys) data set along with 32 active inhibitors with experimental K_i values taken from Tikhe et al.^{54,55} Schrödinger's Glide program was used for all docking studies with the XP scoring function except for where it is mentioned otherwise in the 1F40 study (where AutoDock was used).^{56,57} Hydrogens were added to the crystal structures, and the structures were relaxed with the OPLS 2001 force field within the Maestro program prior to grid generation and docking. These final structures from Maestro (receptors and ligand poses) were used in the AM1 single-point calculations to obtain atomic Fukui indices.

3.1.1. Ranking Ligand Poses in a Receptor. The first docking test was to determine the correct pose of a ligand in the active site of a receptor. The ligand from the FKBP12 system was docked to FKBP12 with the Autodock program.⁵⁸ Ten of the best poses from the docking results were taken to evaluate the hardness and softness matching between atoms in the docked conformations. A score was developed to measure the complementarity of a given ligand and its receptor, hereafter called the FRMSD, or the root-mean-square difference in the Fukui index. For each atom in the docked ligand (L_i), the nearest atom of the protein (R_i) was matched to it to form a closest match atomic pair. The difference in Fukui indices for each atomic pair is squared and then averaged over all ligand atoms (eq 10). A lower value of FRMSD represents a better ligand pose with respect to the match between the hardness and softness of the atoms in the two molecules.

$$\text{FRMSD} = \sqrt{\frac{\sum_i (f_{L_i} - f_{R_i})^2}{N}} \quad (10)$$

A score was calculated for the 10 best poses generated by Autodock, and the results are shown in Figure 1. The two

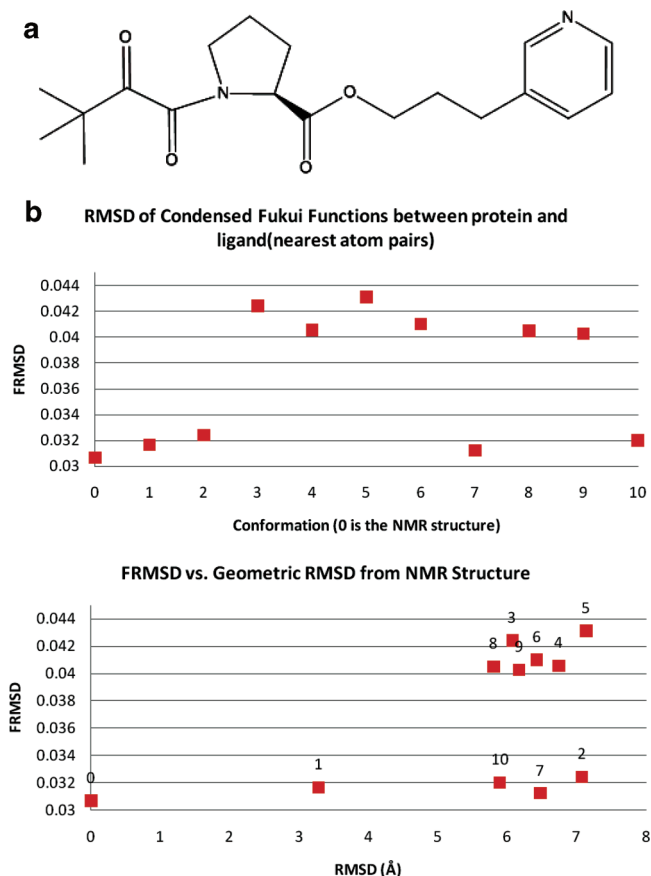


Figure 1. (a) GPI molecule used in the docking procedure for FKBP12. (b, top) Docking of a known binder (GPI-1046) to FKBP12 (PDB ID 1F40). Each point on the horizontal axis represents a different pose taken from the top 10 poses generated by the program Autodock, arranged by decreasing Autodock score. Zero on the horizontal axis represents the observed NMR structure. Low FRMSD values indicate a better hard/soft match between the ligand and its receptor. (b, bottom) FRMSD vs geometric rmsd from the NMR docked structure.

best scoring poses from the Autodock run (poses 1 and 2) score well with the FRMSD score, and worse poses from the Autodock score generally score worse with the FRMSD score. Perhaps the most interesting finding is that the observed pose from the NMR structure (pose 0) has the best FRMSD score, meaning the observed pose is among the docked poses with the best soft/hard matching between closest atom pairs. To show that the NMR pose is actually an acceptable reference pose, an energy minimization was carried out in AMBER for the ligand in the restrained active site. The relaxed ligand structure had an rmsd of 0.255 Å with respect to the NMR ligand structure, which, in our opinion, is a negligible difference.

Figure 1b (bottom) plots the FRMSD of each pose vs the geometric rmsd with respect to the pose from the NMR structure. It was observed that the ligands are divided almost evenly into those with good FRMSD scores and those with poorer FRMSD scores. Upon visualization of the good poses, it was seen that poses 0 and 1 are actually very similar, with the major differences being a rotation of the pyridine ring and a rotation of the *tert*-butyl group. Pose 7 had the same placement of the central pyrrolidine ring but had the positions

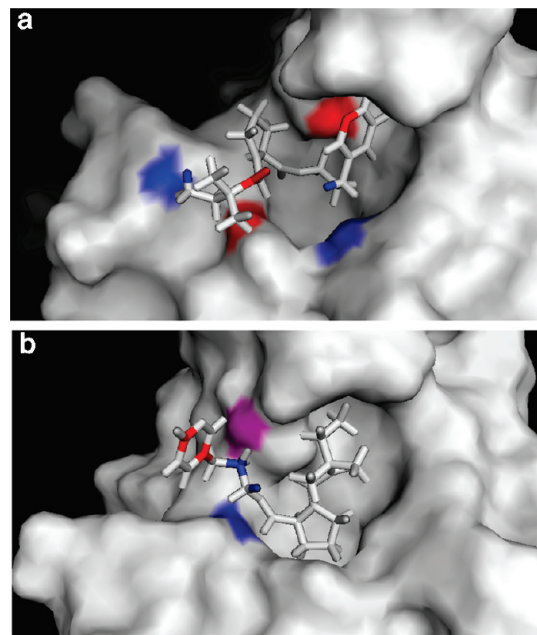


Figure 2. Two docked poses for the FKBP/GPI complex. The active site is shown as a white surface, and the ligand is shown as white sticks. Good hard/soft matching atom pairs are shown in blue and poor hard/soft matches are shown in red on both the ligand and the protein surface. (a) Pose number 5 from the docking procedure. (b) Pose 0, the NMR pose.

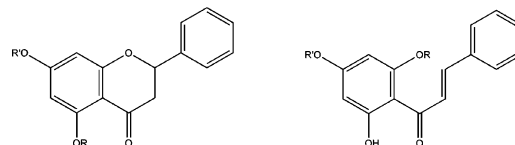


Figure 3. Active binders to dengue type 2 protease taken from a previous docking study by Othman et al.⁵³ (a, left) pinostrobin ($R = H$; $R' = Me$), pinocembrin ($R = H$; $R' = H$), alpinetin ($R = Me$; $R' = H$), (b, right) pinostrobin chalcone ($R = H$; $R' = Me$), pinocembrin chalcone ($R = H$; $R' = H$), cardamonin ($R = Me$; $R' = H$). Reprinted from ref 53. Copyright 2008 American Chemical Society.

of the *tert*-butyl and pyridyl groups swapped (i.e., a molecular rotation by 180°). This pose was also observed in a docking study by Wang et al. in which it was shown to match NMR chemical shift data fairly well.⁵⁹

A benefit of this closest atom pair scoring is that the resulting data can be qualitatively analyzed by simply searching for the best and worst matched pairs. A simple script can analyze the data and produce input for visualization programs such as PyMOL, as demonstrated in Figure 2. Such visual and qualitative measurement of hard/soft matching could be useful in the drug design process, as the human eye can easily detect the best and worst hard/soft matches. In addition, it provides a method of verifying the FRMSD results. In Figure 2, good contacts are marked by shades of blue and poor contacts are marked by shades of red. Of course one could show as many contacts as desired, but here only the two best matches and the two worst matches are shown.

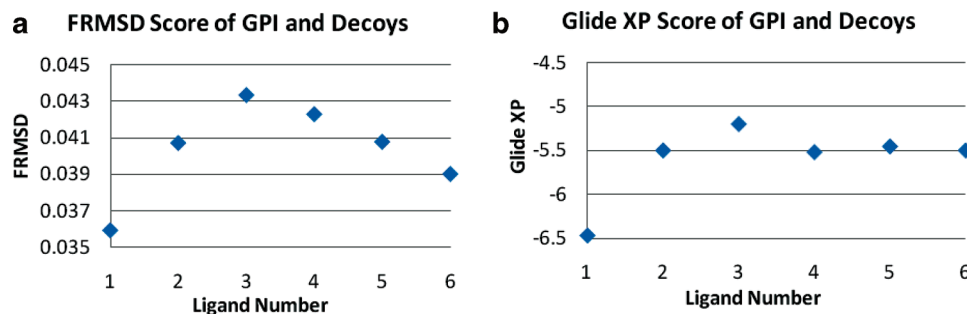


Figure 4. (a) FRMSD scores of the best scoring poses of the GPI molecule and decoy ligands to FKBP12. (b) Glide XP scores. The ligands are numbered as (1) GPI-1046, (2) alpinetin, (3) pinocembrin, (4) pinocembrine chalcone, (5) pinostrobin, and (6) pinostrobin chalcone. Both scores correctly discriminate the known binder from the decoy ligands.

Pose number 5 from the docking run had the worst FRMSD score out of all 10 poses, and its best and worst pairs are shown in Figure 2a. This figure highlights an important interaction not contained in the Fukui function—hydrogen bonding. One of the worst hard/soft mismatches is between the pyridyl nitrogen and the hydroxyl hydrogen from a tyrosine residue inside the binding pocket, which are at a distance of 2.26 Å from each other. This should be a favorable interaction, but is considered a poor interaction from a hard/soft perspective. This example suggests that the Fukui function alone would not be able to account for all types of molecular interactions and would need contributions from additional terms in a scoring function (such as an electrostatic or hydrogen-bonding term) to be universally applicable or to correctly predict binding affinity. In the meantime it is assumed that ligands of similar construction with similar types of interactions can be analyzed by hardness and softness alone.

The NMR pose is shown in Figure 2b. The native pose has its best contacts just outside the binding pocket, and the worst contacts are with the terminal pyridyl group, which faces outward from the binding site. A tyrosine residue near the pocket is shaded purple because it makes both good contacts with the carbon chain of the ligand and poor contacts with the pyridyl group. The color-coding helps in qualitatively understanding why the native pose is a good docking pose. The pyridyl group may have poor hard/soft matching with the receptor, but it is directed outward from the binding site, making the interaction longer ranged and possibly less unfavorable. This would suggest that if distance were taken into account in the scoring function, the native pose would be even more preferred by FRMSD. From visualizing this pose it seems that another way to improve the FRMSD score would be to include a distance dependence, which is introduced in section 3.1.3.

3.1.2. Selection of Active Ligands from Decoys. The second docking test involved the same receptor (FKBP12) with its active binder, GPI-1046, along with a set of decoy ligands from the data set for the dengue virus protease shown in Figure 3. Though these are known binders for the dengue virus type 2 protease, they are assumed to be nonbinding decoys for the FKBP12 system. The top 10 docked poses of each ligand from Glide XP were retained and scored with the closest atom pair FRMSD score. Figure 4a shows only the FRMSD scores for the best scoring pose of each ligand.

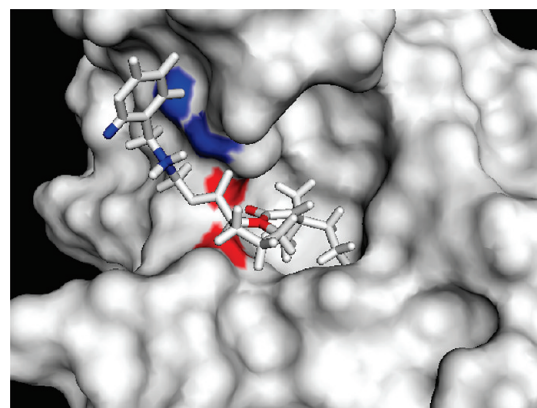


Figure 5. Worst of 10 poses of the correct ligand in the binding site of FKBP12 as determined by the FRMSD score. The poorest hard/soft matches (shown in red) are closer contacts than the good matches (shown in blue), showing that this pose is poorly docked according to hard/soft matching.

Figure 4b shows the Glide XP scores of the best scoring poses for comparison.

Both FRMSD and Glide XP were able to score the correct ligand, GPI-1046, as the best binder. In fact, almost all 10 of the poses generated by Glide scored better than all of the decoy poses. Upon visualization of the worst FRMSD scoring pose of GPI-1046 (Figure 5), the poorest matching pairs are a carbonyl oxygen in the ligand with a γ -methyl hydrogen from an isoleucine residue (at 2.74 Å) and the nitrogen from the ligand's pyrrolidine ring with the α -hydrogen of a valine residue (at 3.66 Å). Both of these pairs should represent somewhat favorable electrostatic interactions, which are not captured by the FRMSD score. The best pairs are between the ligand and a tyrosine group, but both of these pairs are at a distance greater than 3.0 Å, leading to a match that is probably overaccounted for in the distance-independent FRMSD score. This pose provides more evidence that distance should be accounted for in a score based on Fukui indices.

3.1.3. Ranking of Different Ligands by Binding Affinity. The third type of docking experiment for hardness/softness-based scoring was to rank ligand molecules by binding affinity using the Fukui indices for the ligands and receptor. From visualization of the previous docking results it is apparent that a distant-dependent score is necessary. Here a second score is introduced, hereafter referred to as the Fukui grid score, in which distance dependence to the Fukui indices

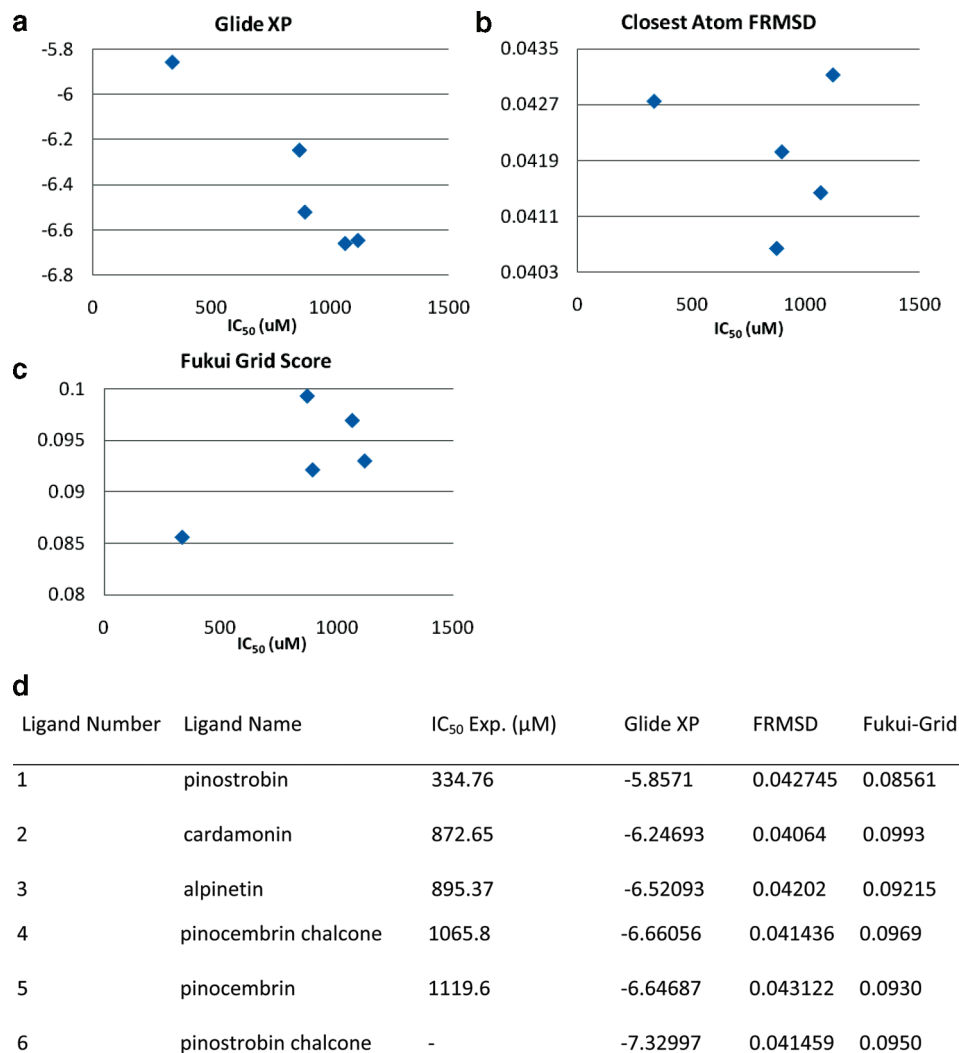


Figure 6. Docking and scoring results from known binders to dengue virus type 2 protease. The Glide XP (a), closest atom pair FRMSD (b), and Fukui grid (c) scores are presented along with the experimental IC₅₀ values (d). The plotted scores represent the pose that yielded the best score for the particular scoring function.

is included. As mentioned before, atomic Fukui indices approximate the full Fukui function as a collection of points centered at atomic positions, which introduces error by ignoring a substantial amount of information about the topology of the Fukui function. In addition, the closest atom pair approach would not properly account for hard/soft matches or mismatches between functional groups. A distance-dependent score could reduce errors caused by both of these factors by allowing many atomic indices to contribute to a value for a given point in space.

The score is calculated by placing a grid of points over each conformation of each ligand. At each grid point, all atomic Fukui indices are first scaled according to their distance to the grid point and then summed. A grid of values is calculated for the receptor's active site as well as for each ligand pose. The grid of each ligand pose is then compared to the receptor grid to generate a score based on eq 11. The grids are superimposed, and each overlapping grid point is used to produce an rmsd between grids. A lower score implies a better match between the hard and soft areas of the receptor and ligand grids. The grids used here were cubic with each side 10 Å in length.

Grid points were spaced by 1.0 Å. Since the distance dependence was unknown, the indices were divided by distance raised to the α power. The parameter α was varied to find the best discrimination between ligands, and for this case a value of $\alpha = 0.5$ was found to be appropriate.

$$\text{grid score} = \sqrt{\frac{\sum_k^{N_{\text{GP}}} [\text{GP}_{\text{ligand}} - \text{GP}_{\text{receptor}}]^2}{\text{number of grid points}}}$$

$$\text{GP}_i = \sum_j \left(\frac{f_j}{r_{jk}^\alpha} \right) \quad (11)$$

The first data set tested was the dengue virus type 2 protease (PDB ID 2FOM) and a collection of known binders that has been previously reported along with IC₅₀ values (Figure 3).^{52,53} The ligands are known to be allosteric binders, as discussed by Othman et al.⁵³ The same binding site was used in this study as was used previously, but whereas Othman used the standard Glide score, here the Glide XP score is used. The top 10 poses of each ligand

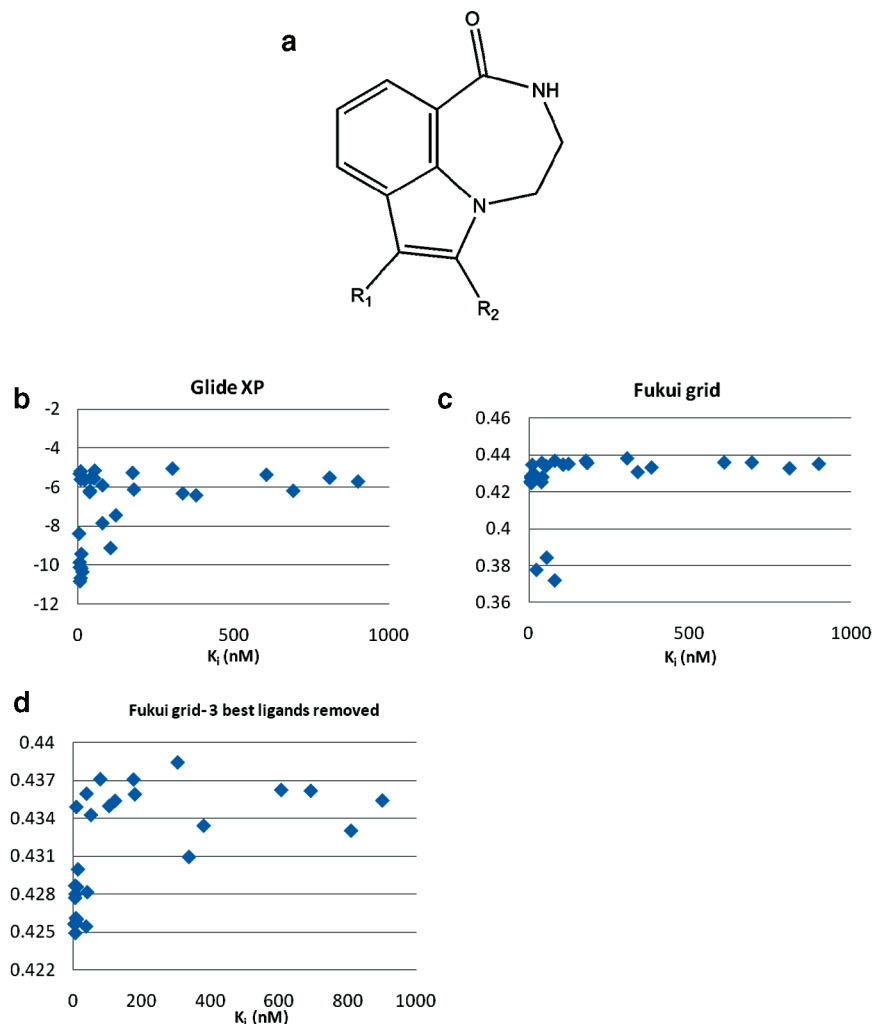


Figure 7. (a) Shared molecular scaffold of the 32 ligands from the DUD data set for the poly(ADP-ribose) polymerase receptor. (b) Results from Glide XP. (c) Fukui grid score. (d). Fukui grid score with the three best ligands removed to display the smaller differences between the remaining ligands.

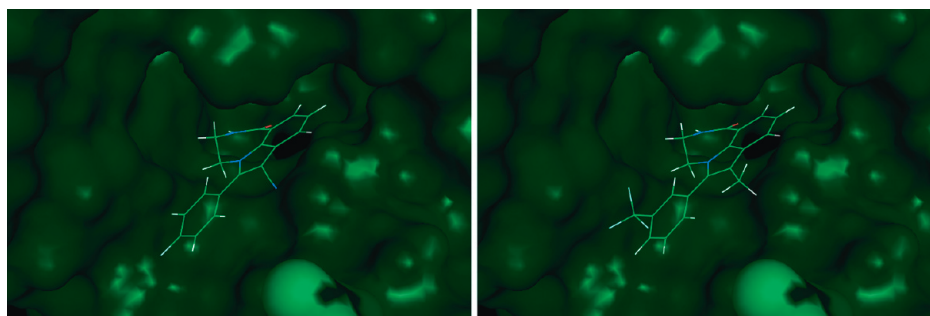


Figure 8. Trifluoromethyl-containing ligand (a) and cyano-containing ligand (b) in their docked poses. The ligand atoms are colored by atom name and are shown as sticks. The protein receptor is shown as a surface that is colored by its atomic Fukui indices. Dark green corresponds to lower values of the Fukui function, and bright colored areas are local maxima in the Fukui function. The relatively hard trifluoromethyl group in (a) points away from the binding site, and the relatively soft cyano group of (b) points toward a soft area just outside the binding pocket.

were saved and scored by their Glide XP score, the closest atom pair FRMSD, and the Fukui grid score. For each scoring function, the best scoring pose from the 10 poses was used to rank the ligands. The results are shown in Figure 6.

The Glide XP score predicts ligand 6, pinostrobin chalcone, to be the best binder and ligand 1, pinostrobin (the

best binder experimentally), to be the worst. The score fails to predict the correct trend in binding affinity and actually predicts the reverse trend. The FRMSD score also fails to rank the ligands correctly or show any kind of trend in binding affinity. In contrast, the distance-dependent Fukui grid score captures the correct trend in binding affinity. It correctly predicts pinostrobin to be the best binder, and the

Table 1. Results from the Active Site Search Experiment^a

| receptor (PDB ID) | active site residue no. | percentile rank |
|-------------------|-------------------------|-----------------|
| Ampc (1XGJ) | 317 | 93.7 |
| Ar (1XQ2) | 873 | 93.0 |
| Fgfr1 (1AGW) | 512 | 97.5 |
| | 545 | 92.4 |
| | 563 | 91.4 |
| | 567 | 92.1 |
| Fxa (1F0R) | 98 | 92.3 |
| | 220 | 98.7 |

^a Four of the twelve studied enzymes had residues within 7 Å of the active site with average atomic Fukui indices ranking higher than those of 90% of the total number of residues in the enzyme.

others are separated from it by hard/soft compatibility. Experimentally this is the case: pinostrobin is by far the strongest binder, while the other ligands are grouped together at lower activity. In this case, hardness/softness matching is able to pick out the best binder from a set of ligands for dengue 2 NS3 protease.

A second test of the Fukui grid score was to rank a set of known binders for poly(ADP-ribose) polymerase (PDB ID 1EFY) taken from the DUD database.^{54,55} The ligands are all similar in structure, with substitutions made at both the R₁ and R₂ positions of the molecular scaffold shown in Figure 7a. The grid score was used as it was in the case for 2FOM, except here a cutoff radius is introduced for the protein Fukui indices due to the larger size of 1EFY. The cutoff used was 10 Å. Increasing this cutoff radius changed individual scores for ligands but did not alter the relative rankings of the ligands. Results from the docking of the 32 DUD ligands are shown in Figure 7.

Here the Glide XP score does fairly well at showing a correct trend in experimental binding affinity (Figure 7b). The Fukui grid score shows only small differences in affinity for the different ligands except for three which clearly display better hard/soft contacts with the receptor than the rest of the set (Figure 7c). If those three are removed from the plot, upon closer examination (Figure 7d) the Fukui grid score still shows better hard/soft matches for many of the stronger binders ($K_i < 50$ nM) than the weaker binders ($K_i > 50$ nM). There were three ligands with K_i less than 60 nM that had relatively poor Fukui grid scores. Two of these ligands had oxime groups participating in hydrogen bonding, and the third had two hydroxyl groups both participating in hydrogen

bonding. As stated previously, the Fukui function based score does not include this stabilizing interaction and would need an additional hydrogen-bonding term to be used as a more general scoring function.

Upon examination of two of the three that stood out as the best hard/soft matching poses, it was found that the best pose evaluated by the Fukui grid score was the only ligand in the DUD set with a trifluoromethyl group (more specifically, 3-(trifluoromethyl)phenyl) which was placed in the R₂ position. Fluorine is usually a very hard atom, so it can be hypothesized that the active site favors a chemically hard group in the area where the trifluoromethyl group is docked. Another of the three best ligands was the only one containing a cyano group in the R₁ position. Cyano groups are usually chemically soft, so the receptor may favor soft species in this position. After visualization of the docked poses of these two ligands, it was discovered that the best poses according to the Fukui grid score had nearly identical binding conformations, with the trifluoromethyl and cyano groups of the two ligands pointing in opposite directions (Figure 8). It would be interesting to test the binding affinity of a ligand containing both of these groups in the two positions to see whether the measured affinity increases due to HSAB preferences. Such a ligand was built and placed in the bound conformation of Figure 8 and then scored by the Fukui grid score. The hypothetical ligand received a score of 0.379, which places it among the best three known ligand molecules in terms of hard/soft compatibility.

3.2. Active Site Detection. The Fukui function is generally used to detect favorable sites of interaction between molecules. Maxima in the function are interpreted as areas in a molecule most favorable for changes in electron density. One can hypothesize that, by mapping a Fukui function, one could easily pick out reactive areas in a molecule. In the realm of proteins, these are called active sites. Fukushima et al. have previously studied the link between the locations of active site residues and localized frontier orbitals.⁶⁰ Among their results was that, for the 112 enzymes under their study, about 20% of the active site residues had molecular orbitals localized on them that lay within a spread of 10 molecular orbitals around the HOMO–LUMO gap.

To use the Fukui function to find active sites of proteins under the finite difference approximation, it is useful to take

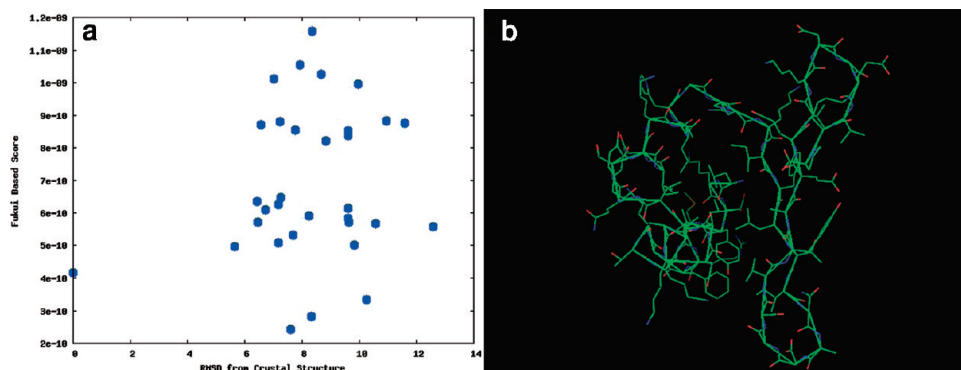


Figure 9. (a) Fukui index based folding score (eq 13) plotted against the rmsd from the native structure for the cro repressor mutant (PDB ID 1ORC) and its decoy folds. The decoy structures have a wide range of scores, and the native structure is among the best scoring folds. (b) Crystal structure of 1ORC.

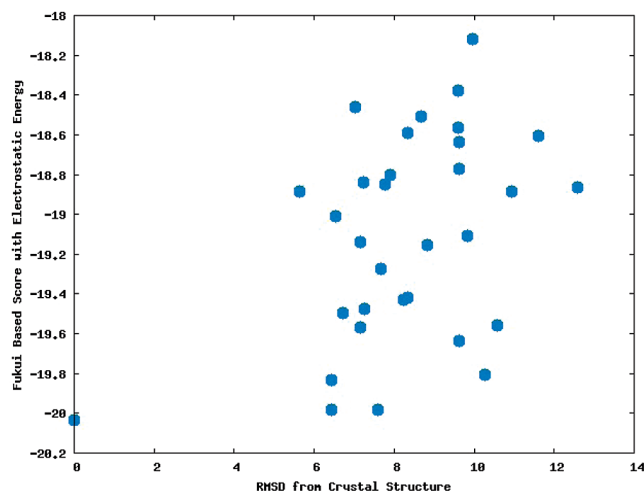


Figure 10. Hybrid score (eq 14) composed of the Fukui function based score (eq 13) and the electrostatic energy score (eq 15) vs the rmsd with respect to the native structure. The native fold scored the best in terms of hard and soft matching, but is not clearly distinguished from the decoy set.

the finite difference derivative by varying the number of electrons by more than 1. Here the electron number was varied by 8. Increasing this number is analogous to increasing the span of MOs searched in Fukushima's study. A calculation was performed on the ground-state system with 8 electrons added and the ground-state system with 8 electrons removed. The centered finite difference Fukui function is then

$$f(r)^0 = \frac{\rho(r, N+8) - \rho(r, N-8)}{2 \times 8} \quad (12)$$

Twelve receptors taken from the DUD database were used in this study.⁵⁴ The experimentally observed bound structures were examined, and all residues within 7 Å of the bound ligand were considered to be part of the active site. The atomic Fukui indices were averaged on each residue to yield one characteristic value of the Fukui function for each amino acid. Each amino acid was then sorted by average Fukui index, and the active site residues were examined by a percentile rank (e.g., 90% means a Fukui score higher than that of 90% of the total number of residues in the enzyme). Among the 12 receptors, 4 of them had active site residues with percentile rankings higher than 90% (Table 1). One of them, fibroblast growth factor receptor 1 (PDB ID 1AGW), had four active site residues ranking higher than 90% of the total number of residues in the protein. This would suggest that polarization of electron density is important for binding to this receptor, which of course is a characteristic that the Fukui function is designed to detect.

If we were to assume these 12 receptors were a randomly chosen collection, then it would make sense that this Fukui function based approach had a success rate similar to that of the Fukushima study. The Fukui function is by definition quite similar to the frontier molecular orbital type analysis used in their earlier work, which is analogous to using the frozen orbital approximation in a Fukui function based approach.

3.3. Protein Folding. The third application explored was the detection of native protein folds from a collection of decoy folds. It was hypothesized that better protein folds should have more favorable hard and soft interactions between residues than poorly folded proteins. To test this hypothesis, a distance-dependent score was introduced, in which Fukui indices of atoms in different amino acids are compared. The score can be written as

$$\sum_i^{N_a} \left(\frac{1}{N_n} \right) \times \begin{cases} \sum_j^{N_n} \left(\frac{f_i - f_j}{r_{ij}^\alpha} \right) & j \notin i_{\text{res}}, r_{ij} < r_{\text{max}} \\ 0 & \text{otherwise} \end{cases} \quad (13)$$

where N_a is the total number of atoms and N_n is the number of neighboring atoms from other residues within a cutoff distance r_{max} . This score was used to rank the X-ray structure and decoy structures of a mutated cro repressor (PDB ID 1ORC). The decoy folds were the same set used by He et al. generated with Rosetta.^{61,62} The values of r_{max} and α were optimized and found to make the best predictions at $r_{\text{max}} = 10$ Å. Varying α did not seem to have a significant impact on the rankings between folds. Figure 9 plots the scores of the X-ray structure and decoys of 1ORC against their rmsd from the X-ray structure. Here $r_{\text{max}} = 10$ Å and $\alpha = 0.2$. Only three of the decoys score better than the native fold, and several of the decoys are far separated from the native structure by the Fukui based score.

As shown in Figure 9, although the Fukui function based score gave the native fold one of the best scores, it could not discriminate the native structure from the set of decoys. As discussed in the docking studies, the hardness and softness interactions do not seem to include electrostatic interactions, which are relevant in looking at protein decoys. Therefore, an electrostatic energy term was added to the Fukui function based score to create a hybrid score given by

$$\text{hybrid score} = x(\text{FF score}) + E_{\text{el}} \text{ score} \quad (14)$$

where x is a parameter introduced to scale the Fukui function based scores (FF score) to the range of the electrostatic energy scores (E_{el} score). The electrostatic energy score used was simply

$$E_{\text{el}} = \sum_{i=1}^{N_a} \sum_{j=i+1}^{N_a} \frac{q_i q_j}{r_{ij}} \quad (15)$$

where q_i is the Mulliken charge of atom i from the AM1 calculation, N_a the number of atoms, and r_{ij} the distance between atoms. An appropriate value for x was found to be around 1.2×10^9 . While the hybrid score was able to rank the native structure the best of all folds, it did not offer clearly superior discrimination ability (Figure 10).

The Fukui function based folding score was also tested on NMR structures of the Pin 1 WW domain (PDB ID 1I6C). This system was chosen because the effect of electron correlation (and attractive van der Waals energy) has been shown to be vital in determining its native fold from a set of decoys.⁶¹ Of the possible nonbonded interactions of molecules, chemical softness seems to be the most relevant to these types of interactions. Using the same parameters as

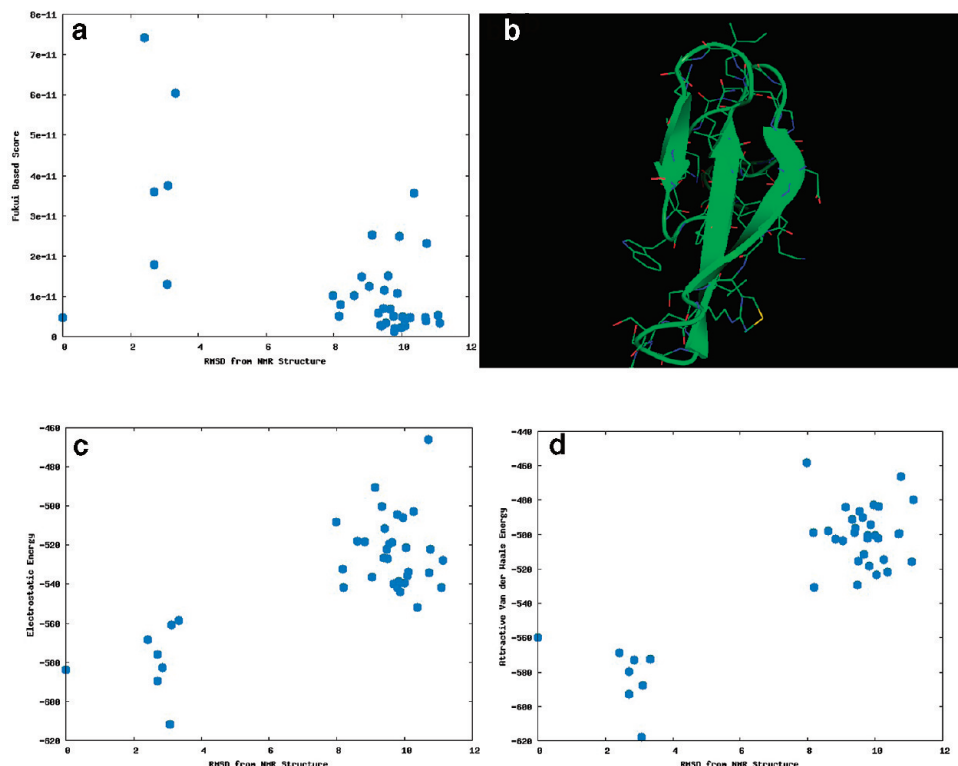


Figure 11. (a) Fukui based folding score for the Pin 1 WW domain (PDB ID 1I6C) vs the rmsd of an NMR structure. The score fails to clearly distinguish the NMR models from the decoys, presumably because the β -sheet conformation keeps the side chains apart (b). (c) Electronic energy from Amber vs the rmsd from the NMR structure. (d) Attractive van der Waals energy from Amber vs rmsd.

with the IORC system, the results of the Fukui based score did not distinguish the NMR structures from the decoys (Figure 11a). Using a hybrid score such as eq 14 would be meaningless in this situation since electronic energy (Figure 11c) or van der Waals energy (Figure 11d) alone do very well in discriminating native from decoy folds. The optimum weighting factor for the Fukui based score would be very close to zero.

One possible source of failure for the Fukui based score would be that the Pin 1 WW domain consists mostly of a β -sheet formation, which is held together by a chain of hydrogen bonds and points side chains outward away from one another. In this kind of conformation, there is little chance for side chains to interact in a hard/soft type of interaction. In contrast, IORC (Figure 9b) has side chains pointed inward near each other. This suggests that any kind of hard/soft scoring for protein folding would most likely be only useful for specific types of protein folds with well-defined cores.

4. Conclusions

The utility of the Fukui function was explored in three significant problems in computational biochemistry: docking, active site detection, and protein folding. We hypothesized that hard/soft acid–base matching concepts would allow us to gain new insights into these problems. To make use of the Fukui function for large protein systems, several approximations were used including the AM1 Hamiltonian, the divide and conquer algorithm, Mulliken charges, and the

finite difference derivative. Even with these approximations Fukui based scoring functions correctly determined the binding conformation of a ligand in an active site, distinguished between active binders and nonbinders for a receptor, determined the best binders from a set of known binders, detected possible active sites, and ranked an observed protein fold among the best of a set of native and decoy folds.

It was observed that not all types of molecular interactions are captured by Fukui based scoring functions and that additional terms (such as an electrostatic term) are necessary to make them more broadly applicable. It was also observed that strictly using atomic indices is not always as effective as approximating the full Fukui function by adding distance dependence, especially in the case of docking several different ligands to a binding site. A clear advantage of these types of analyses is that molecular surfaces can easily be visualized and colored by hardness or softness, aiding the chemist in deciding which parts of two molecules interact favorably or unfavorably from a hardness/softness perspective. The concepts presented herein offer new descriptors that can be used in QSAR studies and present alternative ways to examine biological problems such as protein–ligand interactions.

Acknowledgment. We thank the NIH for generously supporting this research via Grant GM044974.

References

- (1) Szabo, A.; Ostlund, N. S. The Hartree Fock Approximation. *Modern Quantum Chemistry: Introduction to Advanced*

- Electronic Structure Theory*, 1st ed.; Dover Publications: Mineola, NY, 1996; Vol. 1, pp 108–151.
- (2) Levine, I. N. *Ab Initio and Density Functional Treatments of Molecules. Quantum Chemistry*, 6th ed.; Prentice Hall: Upper Saddle River, NJ, 2008; Vol. 1, pp 480–625.
 - (3) Jensen, F. *Electronic Structure Methods: Independent-Particle Models. Introduction to Computational Chemistry*, 2nd ed.; Wiley: Chichester, England, 2006; Vol. 1, pp 80–132.
 - (4) Cramer, C. J. *Molecular Mechanics. Essentials of Computational Chemistry: Theories and Models*, 2nd ed.; Wiley: Chichester, England, 2004; Vol. 1, pp 17–68.
 - (5) Leach, A. *Empirical Force Field Models: Molecular Mechanics. Molecular Modelling: Principles and Applications*, 2nd ed.; Prentice Hall: Harlow, England, 2001; Vol. 1, pp 165–252.
 - (6) Geerlings, P.; De Proft, F.; Langenaeker, W. *Chem. Rev.* **2003**, *103*, 1793.
 - (7) Sengupta, D.; Chandra, A. K.; Nguyen, M. T. *J. Org. Chem.* **1997**, *62*, 6404.
 - (8) Roy, R. K.; Krishnamurti, S.; Geerlings, P.; Pal, S. *J. Phys. Chem. A* **1998**, *102*, 3746.
 - (9) Roy, R. K.; Tajima, N.; Hirao, K. *J. Phys. Chem. A* **2001**, *105*, 2117.
 - (10) Feng, X.-T.; Yu, J.-G.; Lei, M.; Fang, W.-H.; Liu, S. *J. Phys. Chem. B* **2009**, *113*, 13381.
 - (11) Beck, M. E. *J. Chem. Inf. Model.* **2005**, *45*, 273.
 - (12) Roos, G.; Geerlings, P.; Messens, J. *J. Phys. Chem. B* **2009**, *113*, 13465.
 - (13) Giessner, C.; Pullman, A. *Theor. Chim. Acta* **1972**, *25*, 83.
 - (14) Besler, B. H.; Merz, K. M.; Kollman, P. A. *J. Comput. Chem.* **1990**, *11*, 431.
 - (15) Goedecker, S. *Rev. Mod. Phys.* **1999**, *71*, 1085.
 - (16) Galli, G. *Phys. Status Solidi B* **2000**, *217*, 231.
 - (17) Dixon, S. L.; Merz, K. M. *J. Chem. Phys.* **1996**, *104*, 6643.
 - (18) Dixon, S. L.; Merz, K. M. *J. Chem. Phys.* **1997**, *107*, 879.
 - (19) Van der Vaart, A.; Gogonea, V.; Dixon, S. L.; Merz, K. M. *J. Comput. Chem.* **2000**, *21*, 1494.
 - (20) Yang, W. T.; Lee, T. S. *J. Chem. Phys.* **1995**, *103*, 5674.
 - (21) Lee, T. S.; Lewis, J. P.; Yang, W. T. *Comput. Mater. Sci.* **1998**, *12*, 259.
 - (22) Khandogin, J.; York, D. M. *Proteins* **2004**, *56*, 724.
 - (23) Pearson, R. G. *J. Am. Chem. Soc.* **1963**, *85*, 3533.
 - (24) Datta, D.; Singh, S. N. *J. Chem. Soc., Dalton Trans.* **1991**, 1541.
 - (25) Datta, D. *J. Chem. Soc., Dalton Trans.* **1992**, 1855.
 - (26) Benedetti, L.; Gavioli, G. B.; Fontanesi, C. *J. Chem. Soc., Faraday Trans.* **1992**, *88*, 843.
 - (27) Langenaeker, W.; Coussement, N.; Deproft, F.; Geerlings, P. *J. Phys. Chem.* **1994**, *98*, 3010.
 - (28) Deproft, F.; Amira, S.; Choho, K.; Geerlings, P. *J. Phys. Chem.* **1994**, *98*, 5227.
 - (29) Deka, R. C.; Vetrivel, R.; Pal, S. *J. Phys. Chem. A* **1999**, *103*, 5978.
 - (30) Mondal, P.; Hazarika, K. K.; Deka, R. C. *PhysChemComm* **2003**, *24*.
 - (31) Flores-Sandoval, C. A.; Zaragoza, I. P.; Maranon-Ruiz, V. F.; Correa-Basurto, J.; Trujillo-Ferrara, J. *J. Mol. Struct.: THEOCHEM* **2005**, *713*, 127.
 - (32) Wisniewski, M.; Gauden, P. A. *Appl. Surf. Sci.* **2009**, *255*, 4782.
 - (33) Parr, R. G.; Pearson, R. G. *J. Am. Chem. Soc.* **1983**, *105*, 7512.
 - (34) Parr, R. G.; Yang, W. T. *J. Am. Chem. Soc.* **1984**, *106*, 4049.
 - (35) Yang, W. T.; Parr, R. G. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 6723.
 - (36) Parr, R. G.; Donnelly, R. A.; Levy, M.; Palke, W. E. *J. Chem. Phys.* **1978**, *68*, 3801.
 - (37) Ayers, P. W.; Parr, R. G. *J. Chem. Phys.* **2008**, *128*, 184108.
 - (38) Chattaraj, P. K.; Roy, D. R.; Geerlings, P.; Torrent-Sucarrat, M. *Theor. Chem. Acc.* **2007**, *118*, 923.
 - (39) Ayers, P. W.; De Proft, F.; Borgoo, A.; Geerlings, P. *J. Chem. Phys.* **2007**, *126*, 224107.
 - (40) Fievez, T.; Sablon, N.; De Proft, F.; Ayers, P. W.; Geerlings, P. *J. Chem. Theory Comput.* **2008**, *4*, 1065.
 - (41) Chandra, A. K.; Nguyen, M. T. *J. Chem. Soc., Faraday Discuss.* **2007**, *135*, 191.
 - (42) Balawender, R.; Komorowski, L. *J. Chem. Phys.* **1998**, *109*, 5203.
 - (43) Mineva, T.; Russo, N.; Sicilia, E.; Toscano, M. *Theor. Chem. Acc.* **1999**, *101*, 388.
 - (44) Mineva, T.; Parvanov, V.; Petrov, I.; Neshev, N.; Russo, N. *J. Phys. Chem. A* **2001**, *105*, 1959.
 - (45) Madjarova, G.; Tadjer, A.; Cholakova, T. P.; Dobrev, A. A.; Mineva, T. *J. Phys. Chem. A* **2005**, *109*, 387.
 - (46) Ayers, P. W. *Phys. Chem. Chem. Phys.* **2006**, *8*, 3387.
 - (47) Melin, J.; Ayers, P. W.; Ortiz, J. V. *J. Phys. Chem. A* **2007**, *111*, 10017.
 - (48) Vincent, J. J.; Dixon, S. L.; Merz, K. M. *Theor. Chem. Acc.* **1998**, *99*, 220.
 - (49) DeLano, W. L.; Lam, J. W. *Abstr. Pap.—Am. Chem. Soc.* **2005**, *230*, U1371.
 - (50) Sich, C.; Improta, S.; Cowley, D. J.; Guenet, C.; Merly, J. P.; Teufel, M.; Saudek, V. *Eur. J. Biochem.* **2000**, *267*, 5342.
 - (51) Erbel, P.; Schiering, N.; D'Arcy, A.; Renatus, M.; Kroemer, M.; Lim, S. P.; Yin, Z.; Keller, T. H.; Vasudevan, S. G.; Hommel, U. *Nat. Struct. Mol. Biol.* **2006**, *13*, 372.
 - (52) Kiat, T. S.; Phippen, R.; Yusof, R.; Ibrahim, H.; Norzulaani, K.; Rahman, N. A. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 3337.
 - (53) Othman, R.; Kiat, T. S.; Khalid, N.; Yusof, R.; Newhouse, E. I.; Newhouse, J. S.; Alam, M.; Rahman, N. A. *J. Chem. Inf. Model.* **2008**, *48*, 1582.
 - (54) Huang, N.; Shoichet, B. K.; Irwin, J. J. *J. Med. Chem.* **2006**, *49*, 6789.
 - (55) Tikhe, J. G.; Webber, S. E.; Hostomsky, Z.; Maegley, K. A.; Ekkers, A.; Li, J. K.; Yu, X. H.; Almassy, R. J.; Kumpf, R. A.; Boritzki, T. J.; Zhang, C.; Calabrese, C. R.; Curtin, N. J.; Kyle, S.; Thomas, H. D.; Weng, L. Z.; Calvert, A. H.; Golding, B. T.; Griffin, R. J.; Newell, D. R. *J. Med. Chem.* **2004**, *47*, 5467.
 - (56) Halgren, T. A.; Murphy, R. B.; Banks, J.; Mainz, D.; Klicic, J.; Perty, J. K.; Friesner, R. A. *Abstr. Pap.—Am. Chem. Soc.* **2002**, *224*, U345.

- (57) Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. *J. Med. Chem.* **2006**, *49*, 6177.
- (58) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. *J. Comput. Chem.* **1998**, *19*, 1639.
- (59) Wang, B.; Westerhoff, L. M.; Merz, K. M. *J. Med. Chem.* **2007**, *50*, 5128.
- (60) Fukushima, K.; Wada, M.; Sakurai, M. *Proteins* **2008**, *71*, 1940.
- (61) He, X.; Fusti-Molnar, L.; Cui, G. L.; Merz, K. M. *J. Phys. Chem. B* **2009**, *113*, 5290.
- (62) Bonneau, R.; Strauss, C. E. M.; Rohl, C. A.; Chivian, D.; Bradley, P.; Malmstrom, L.; Robertson, T.; Baker, D. *J. Mol. Biol.* **2002**, *322*, 65.

CT9005085