Quantifying Intrinsic Specificity: A Potential Complement to Affinity in Drug Screening

Jin Wang,[1,2,*] Xiliang Zheng,[1] Yongliang Yang,[2] Dale Drueckhammer,[2] Wei Yang,[3] Gennardy Verkhivker,[4] and Erkang Wang[1]

[1]*State Key Laboratory of Electroanalytical Chemistry, Changchun Institute of Applied Chemistry, Chinese Academy of Sciences Changchun, Jilin 130022 People's Republic of China*

[2]*Department of Chemistry and Department of Physics, State University of New York at Stony Brook, Stony Brook, New York 11794-3400, USA*

[3]*Department of Chemistry, Florida State University, Tallahassee, Florida 32306-4390, USA*

[4]*Pfizer Global Research and Development, La Jolla Laboratories, 10777, Science Center Drive, San Diego, California 92121, USA*

(Received 23 October 2006; published 6 November 2007)

We report here the investigation of a novel description of specificity in protein-ligand binding based on energy landscape theory. We define a new term, intrinsic specificity ratio (ISR), which describes the level of discrimination in binding free energies of the native basin for a protein-ligand complex from the weaker binding states of the same ligand. We discuss the relationship between the intrinsic specificity we defined here and the conventional definition of specificity. In a docking study of molecules with the enzyme COX-2, we demonstrate a statistical correspondence between ISR value and geometrical shapes of the small molecules binding to COX-2. We further observe that the known selective (nonselective) inhibitors of COX-2 have higher (lower) ISR values. We suggest that intrinsic specificity ratio may be a useful new criterion and a complement to affinity in drug screening and in searching for potential drug lead compounds.

DOI: 10.1103/PhysRevLett.99.198101 PACS numbers: 87.15.−v

Studying biomolecular recognition is critical in understanding the fundamental biological metabolism and signaling events and is also at the core of drug design [1,2]. The two crucial issues related to the binding process are the affinity of the two molecules for each other and the specificity or tendency of a molecule to bind to its desired target instead of other biomolecules. High affinity is often used as the initial criterion in the screening of drug targets in the pharmaceutical industry. However, high affinity does not guarantee the high specificity which is critical for drug target discrimination. An important lesson comes from inhibitors of the highly homologous cyclooxygenase (prostaglandin synthase) enzymes COX-1 and COX-2. Inhibition of COX-2 can reduce inflammation and pain (typical COX-2 inhibitors include aspirin and advil) [3–5]. However, nonspecific binding to COX can cause serious side effects, with over 16 500 deaths and 103 000 hospitalizations per year in the U.S. [6].

While affinity is readily defined as the free energy difference between associated and dissociated states, the definition of specificity is less clear. We have investigated a new approach to specificity based on energy landscape theory.

The conventional definition of specificity is the ability of a specific ligand (by ligand here we mean small molecule) to discriminate against different macromolecular receptors [Fig. 1(a)]. To determine the specificity of a specific ligand for a specific receptor, one has to search all the related receptors and find the set with lowest binding free energies sufficiently separated from the rest (to realize the discrimination in population which is exponentially related to the free energy by Boltzmann law), which is often impractical.

An alternate view of specificity is the capability of a particular macromolecular receptor to discriminate between different ligand molecules [Fig. 1(c)]. A new view of specificity addressed in this Letter is the preference for a particular binding state (or mode) or a particular set of binding modes of a ligand to its receptor [Fig. 1(b)] to be much lower in energy than the other weakly binding states. During the process of a ligand binding to its receptor, different intermediate binding states emerge which have

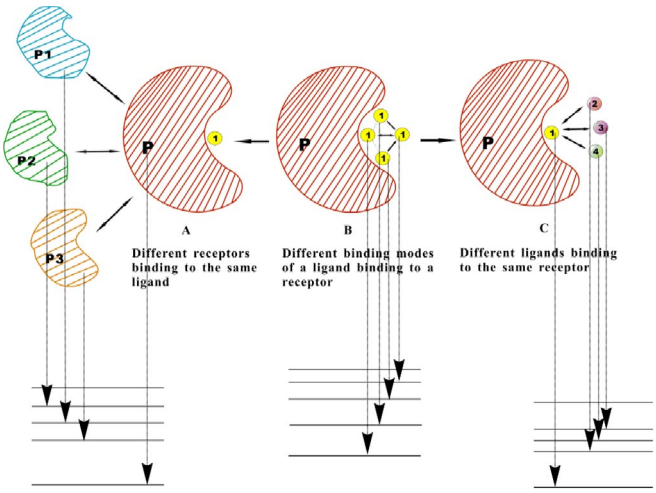

FIG. 1 (color online). Illustration of the concept of specificity in ligand binding and relationship between intrinsic specificity and the conventional specificity as well as the corresponding energy spectrum: (a) Different receptors binding to the same ligand; (b) Different binding states (modes) of a particular ligand to its receptor; (c) Different ligands binding to the same receptor.

different structures with different binding energies and different sets of spatial contact interactions between the ligand and the receptor. Under the ensemble hypothesis in statistical mechanics, if there are sufficient number of spatial contact interactions between the ligand and the receptor, this third view of specificity should be statistically equivalent to the first two in terms of probing the underlying interactions. That is throwing the dice multiple times [a particular ligand in different binding states or modes to its receptor of Fig. 1(b)] is statistically equivalent to throwing multiple dice [multiple receptors binding to the same ligand of Fig. 1(a), or multiple ligands binding to the same receptor of Fig. 1(c)] at once.

Theoreticians have argued that the binding of a ligand to a large protein occurs on a much smaller time scale than would be necessary to search through all the possible configurational states. The binding energy landscape should thus have a funneled shape toward the "native" binding basin as shown in Figure 1 in Ref. [7]) (See also Fig. 2 in this Letter) [7–18]. The bottom of the funnel has a distribution of the native substates as pointed out from the pioneering CO-Myglobin binding experiments of Frauenfelder and his collaborators [19], while the population distribution of the other "non-native" weakly bound states (binding modes) also have a nearly Gaussian distribution. The two important free energy terms are δE (Fig. 2), which represents the energy difference or gap between the native basin (average free energies within the native basin) and "non-native" ones (the average free energy of the "non-native" weakly binding states), and ΔE (Fig. 2), which defines as the square root of the sum of the free energy variance of both native basin and"non-native" weakly binding states. Since usually the variance of the native basin substates are significantly smaller than the "non-native" ones, we can approximate ΔE as $\Delta E = \sqrt{\langle E^2\rangle - \langle E\rangle^2}$ considering only the non-native ones. ΔE has the statistical meaning of half width of the "non-native" binding energy distribution (Fig. 2). The ratio of these two energy terms $\delta E/\Delta E$ is defined as the intrinsic specificity ratio (ISR) (Fig. 2). Intrinsic specificity here means the capability of discriminating the binding states (modes) of the native basin from other "non-native" binding states (modes) for a particular ligand-receptor complex, in contrast to the conventional definition of specificity among different binding partners which could be called extrinsic specificity. Intrinsic specificity introduced here is related to the capability of folding and binding (Z score) [7–14,18]. Since the population follows a Boltzmann distribution, $P \sim \exp[-E/kT]$, a large ISR indicates a high level of discrimination of the free energy states (binding modes) of native basin from the "non-native" weaker binding states (modes) for a particular ligand-receptor complex [7]. Therefore, ISR provides a quantitative measure of intrinsic specificity that can be readily determined without identifying or studying the multitude of alternate receptors or ligands that would have to be evaluated for the conventional "extrinsic" definitions of specificity. Thus, we see the new intrinsic specificity definition has the advantage of being able to quickly identify and quantify the specificity of a ligand to a receptor, without going through all the universe of receptors for a specific ligand (the conventional definition of specificity).

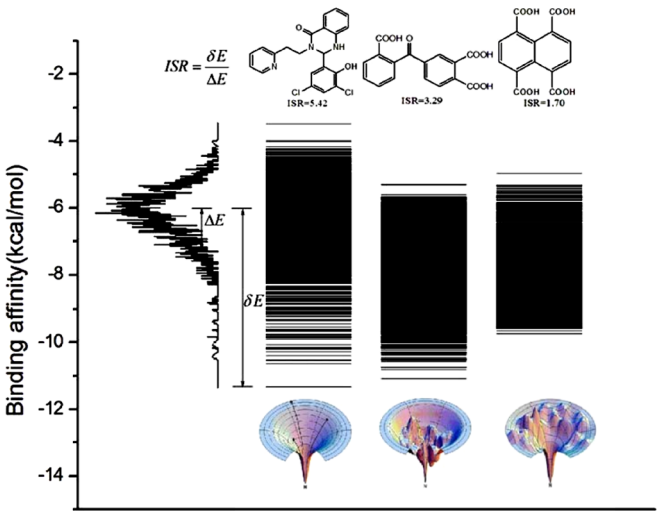

FIG. 2 (color online). The spectrum and the associated distribution of binding energies for resulting high, medium, and low ISR values of three different ligands binding to Cox-2, respectively, and the binding energy funnel for each. Notice that the sparse part of the spectrum implies there are fewer states, and dense part of the spectrum represented by lines implies there are more states. High ISR has a discrimination between the native and non-native basins. For theoretical details, see Ref. [7]. The illustrations of the bottom three funnels are from Ref. [27].

In this Letter, we report the investigation of the significance and implications of ISR using COX-2 as a model. Initially, a diverse set of 1000 small molecules were selected from the National Cancer Institute Diversity database having molecular weights similar to that of the reference compound SC-558, for which the crystal structure of the COX-2 complex is available (PDB code 1CX2) [20]. SC-558 is very similar in structure to three of the COX-2 inhibitory drugs Vioxx, Celebrex, and Bextra, with Celebrex differing from SC-558 only in replacement of the bromine with a methyl group [21].

We performed binding simulations of each small molecule to COX-2. All conformers of each of the 1000 selected molecules were docked with COX-2 using the AUTODOCK program [22] with a standard molecular force field, where the COX-2 protein was held fixed and small molecules were allowed to be flexible. The conformational search of the small molecules was carried out through a local genetic algorithm (LGA) [23]. Enough samples have been used to guarantee the convergence of the final result. The binding free energy is obtained by taking into accounts the Van der Waals forces, hydrogen bonding, electrostatic, desolvation,

and torsions:

$$\Delta G_{\text{binding}} = \Delta G_{\text{vdw}} + \Delta G_{\text{elec}} + \Delta G_{h\text{bond}} + \Delta G_{\text{desolv}} + \Delta G_{\text{tors}}. \quad (1)$$

From this simulation, the binding energy spectrum of all the binding modes and associated ISR for each molecule with COX-2 are generated. From this data, the ISR for each molecule with COX-2 was calculated as $\delta E/\Delta E$. Furthermore, the structures of the small molecules can be classified according to their geometrical shapes, from linear, to V, and to Y. One of the most stable enzyme (COX-2) complexes with a known small molecule drug is SC-558 (often taken as the reference) where the drug molecule has a Y shape.

A plot of intrinsic specificity ratio vs structural shape from linear to V, and to Y, for all the small molecules binding to the same site is shown in Fig. 3. This plot shows a monotonic trend between the two terms, with a statistical correlation coefficient of 0.5. Thus, a higher intrinsic specificity ratio has the tendency to correspond to a more Y like shape for the molecule and therefore a structural match between the small molecule and the reference compound when bound to COX-2. On the other hand, we do not see significant correlation between the affinity and structural shape, which indicates that high affinity does not necessarily correspond to structural consensus.

Figure 2 shows the structures, the plots of the spectrum, and the associated distribution of the binding energies, as well as the underlying binding energy landscape for three representative small molecules binding to COX-2 with high, medium, and low ISR values. The molecule with a high ISR value of 5.42 exhibits a lowest energy native binding basin that is well set apart energetically from the non-native binding states, indicating a steep funnel toward the native state as is expected for a compound exhibiting high intrinsic specificity. This molecule has a Y shape structure essential for selective inhibition of COX-2, while the medium (V shape, ISR = 3.29) and low (linear shape, ISR = 1.70) ISR compounds do not have such a Y structure and thus are expected to be less specific inhibitors of COX-2. These latter two compounds have a smaller energy difference between the native and average non-native states relative to the spread of the energy spectrum of non-native states. The higher ISR compound exhibits a smoother energy landscape while the medium and low ISR compounds have rougher energy landscapes. The composition of the underlying interactions was also investigated. It was observed that hydrophobic interactions [24] dominate in the COX-2 complexes of molecules having high ISR values (over 90% of the total interactions). However, hydrophobic interactions contribute a smaller percentage of the binding with medium (around 80% of the total interactions) and low ISR compounds (around 60% of the total interactions), with other apparently less specific interactions being more important. The origin of high intrinsic specificity thus appears to be the underlying hydrophobic interactions. Since hydrophobic interactions are short range in nature, they are mainly responsible for the shape complementarity between two molecules, especially at the binding interface, and therefore the intrinsic specificity. This may provide a physical origin for the correlation between the specific geometrical shape of the small molecule and the ISR value obtained from the physical binding spectrum.

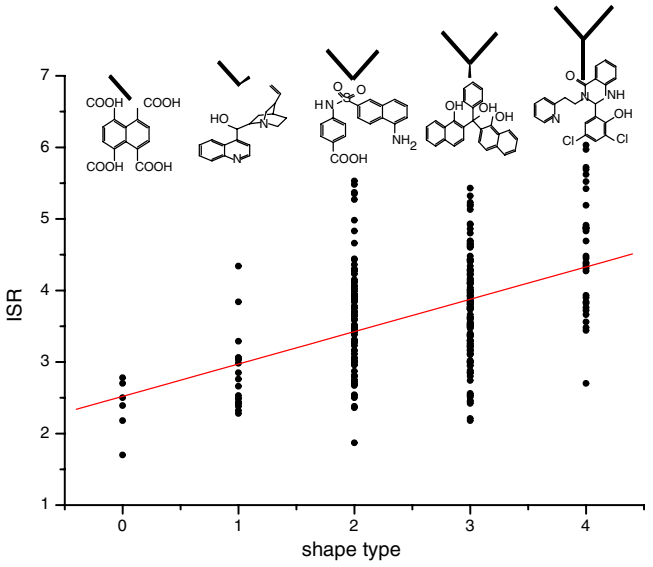

FIG. 3 (color online). Plot of ISR versus geometrical shapes of the small molecules ranging from linear, to V, and to Y binding to COX-2.

In the discovery of novel lead molecules for drug design, initially important criteria are binding affinity and extrinsic specificity for the target protein. Standard screening techniques, whether experimentally or computationally based, generally focus on affinity, which may not correspond to selectivity. The work reported here suggests that in computer-based screening, it may be valuable to evaluate both affinity and intrinsic specificity ratio. Again, using the COX inhibitors as a test system, the known selective inhibitors for COX-2 are observed to have relatively higher ISR values (red dots in Fig. 4) while the nonselective COX inhibitors including the common nonsteroidal anti-inflamatory drugs (NSAIDs) have relatively smaller ISR values [white (transient) and violet (stable) dots in Fig. 4]. The three gray dots correspond to two drugs (rofecoxib and valdecoxib) taken off from the market due to the side effects, and the other one (celecoxib) which has serious problems and is under critical review. When ISR value is roughly beyond 4, all the inhibitors are selective (red dots). Therefore, ISR provides a possible marker for specificity or selectivity. This further suggests that one should look for both high affinity and high ISR value in searching for drug candidates and lead compounds.

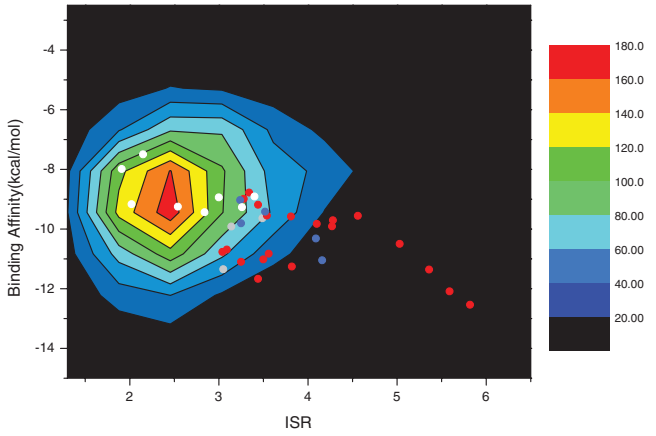

FIG. 4 (color). Contour map of binding affinity and ISR for 1000 small molecules binding with COX-2, including some known COX-2 selective drugs (red dots), classical nonselective NSAIDs (white and violet), and three drugs with serious side effects (gray).

In drug screening, the competing targets that may result in side effects of the drug are not always readily known. It is impractical to experimentally evaluate binding toward all possible competing targets, and structures of potential competing targets are not necessarily available for computational evaluation of affinity. The intrinsic specificity ratio introduced here is relatively easy to calculate and serves as an indicator of the structural match of the target with each small molecule inhibitor. This provides an initial indicator of expected specificity in the absence of information about specific competing targets. This work offers the possibility of two-dimensional computational drug screening using both affinity and intrinsic specificity as selection criteria.

It is worthwhile to point out that quantifying specificity in this work can be helpful for the study of chemical genetics [25,26]. Evolution of the ligands (receptors) can be such that the best ones will match with the specific receptor (ligand) and perform the corresponding biological functions. These ligands (receptors) then would be the ones selected by nature. Using these ligands (receptors) as probes, one can also figure out the biological functions of specific receptor (ligand). (See Figs. 1 and 2.)

This work was supported by National Science Foundation Grant No. CHE0213457 (D. D.) and by a National Science Foundation Career Grant No. 0447533 (J. W.), American Chemical Society Petroleum Research Fund (J. W.), National Science Foundation of China No. 20575063 (E. K. W.), Chinese Academy of Sciences No. KJCX2 (E. K. W.), Special Funds for Major State Basic Research No. 2007CB714500 (E. K. W.). J. W. would like to thank Professor Peter G. Wolynes for an enlightening discussion and insights on the relationship between the intrinsic specificity and the conventional specificity. We would like to thank Prof. Chris Walsh, Professor Francis Arnold, Prof. Zhen-Gang Wang, Professor Yaoqi Zhou, Prof. Jianpeng Ma, Professor Qiang Cui, and Professor Luhua Lai for helpful discussions.

*Corresponding Authors: jin.wang.1@stonybrook.edu.

[1] J. A. McCammon, Curr. Opin. Struct. Biol. **8**, 245 (1998).
[2] A. Wlodawer and J. W. Erickson, Annu. Rev. Biochem. **62**, 543 (1993).
[3] W. L. Smith and D. L. DeWitt, R. M. Garavito, Annu. Rev. Biochem. **69**, 145 (2000).
[4] R. N. Dubois and P. E. Lipsky, FASEB J. **12**, 1063 (1998).
[5] P. E. Lipsky, Am. J. Med. **5**, 515 (1999).
[6] M. M. Wolfe, D. R. Lichtenstein, and G. Singh, New England Journal of Medicine Medical Progress Series **340**, 1888 (1999).
[7] J. Wang and G. M. Verkhivker, Phys. Rev. Lett. **90**, 188101 (2003).
[8] P. A. Rejto and G. M. Verkhivker, Proc. Natl. Acad. Sci. U.S.A. **93**, 8945 (1996).
[9] N. K. Shah, P. A. Rejto, and G. M. Verkhivker, Proteins: structure, function and genetics **28**, 421 (1997).
[10] J. Janin, Proteins **25**, 438 (1996).
[11] M. Vieth *et al.*, J. Comput. Chem. **19**, 1612 (1998).
[12] P. Ferrara *et al.*, J. Med. Chem. **47**, 3032 (2004).
[13] Z. Liu *et al.*, J. Am. Chem. Soc. **126**, 8515 (2004).
[14] B. N. Dominy and E. Shakhnovich, J. Med. Chem. **47**, 4538 (2004).
[15] C. J. Tsai, S. Kumar, B. Ma, and R. Nussinov, Protein Sci. **8**, 1181 (1999).
[16] A. Tovchigrechko and I. A. Vakser, Protein Sci. **10**, 1572 (2001).
[17] P. G. Wolynes, J. N. Onuchic, and D. Thirumalai, Science **267**, 1619 (1995).
[18] A. Goldstein, Z. A. Luthey-Schulten, and P. G. Wolynes, Proc. Natl. Acad. Sci. U.S.A. **89**, 4918 (1992).
[19] H. Frauenfelder, S. G. Sligar, and P. G. Wolynes, Science **254**, 1598 (1991).
[20] R. G. Kurumbail *et al.*, Nature (London) **384**, 644 (1996).
[21] D. L. Dewitt, Molecular Pharmacology **55**, 625 (1999).
[22] D. S. Goodsell, G. M. Morris, and A. J. Olson, J. Mol. Recognit. **9**, 1 (1996).
[23] G. M. Morris and D. S. Goodsell *et al.*, J. Comput. Chem. **19**, 1639 (1998).
[24] Z. Liu and H. S. Chan, J. Mol. Biol. **349**, 872 (2005); S. Shimizu and H. S. Chan, Proteins: structure, function and genetics **48**, 15 (2002).
[25] S. L. Schreiber, Science **300**, 294 (2003).
[26] S. L. Schreiber, Chem. & Eng. News **81**, 51 (2003).
[27] H. S. Chan and K. A. Dill, Proteins: structure, function and genetics **30**, 2 (1998); K. A. Dill and H. S. Chan, Nat. Struct. Biol. **4**, 10 (1997).