Chemical Science

RSC Publishing

EDGE ARTICLE

View Article Online
View Journal | View Issue

Published on 04 April 2013. Downloaded by Pennsylvania State University on 17/09/2016 00:16:25.

Cite this: *Chem. Sci.*, 2013, **4**, 2387

Thermodynamic and kinetic specificities of ligand binding†

Zhiqiang Yan,‡[a] Xiliang Zheng,‡[bc] Erkang Wang[b] and Jin Wang*[ab]

Binding affinity and specificity are crucial for biomolecular recognition. Past studies have focused on binding affinity while the quantification of specificity has remained an elusive challenge. The conventional specificity measures the discrimination of the specific receptor against others for a ligand binding. It is difficult to explore all the possible competing receptors for the ligand. Here, we quantified the thermodynamic intrinsic specificity of discriminating the "native" binding mode against the "non-native" binding modes. Intrinsic specificity is relatively easy to compute since one doesn't need to explore all the other receptors. We found that the thermodynamic intrinsic specificity correlates with the conventional specificity. This validates the statistical equivalence of conventional and intrinsic specificities for receptors at reasonable size. We also computationally quantified the residence time of a ligand on the receptor target as the kinetic specificity. We found that the kinetic specificity correlates with the thermodynamic intrinsic specificity and the binding affinity, suggesting the kinetics and the thermodynamics can be simultaneously optimized for biological activities. With the thermodynamic and kinetic specificities in addition to affinity, we carried out a drug screening test on the target cyclooxygenase-2 (COX-2). We showed that multidimensional (two- and three-dimensional) screening has higher capability than affinity alone to discriminate the drug target (COX-2) from the competitive receptor (COX-1), and the selective drugs from the non-selective drugs. Our work suggests a new way of multidimensional drug screening and target identification, which has significant potential applications for drug discovery and design.

Received 19th February 2013
Accepted 3rd April 2013

DOI: 10.1039/c3sc50478f

www.rsc.org/chemicalscience

1 Introduction

Biomolecular recognition is the basis of all biological processes in living organisms. Understanding of biomolecular recognition at the molecular level is one of the most important problems in modern molecular biology and has direct applications in drug discovery and design.[1–5] Thermodynamically, highly efficient and specific receptor–ligand binding requires two features. One is the affinity related to the stability of the binding. The other is the specificity which discriminates the "good" specific binding complexes from the "bad" non-specific binding ones.[6–11] While tremendous efforts have been devoted into the study of the affinity, the quantitative description of the binding specificity is still challenging.

The conventional thermodynamic specificity concerns about the differences in affinity of the same ligand binding with different receptors[6–11] (Fig. 1A). The high specificity gives the discrimination of one receptor target from others. The concept of conventional thermodynamic specificity is clear, but the actual practice of quantification is challenging. In practice, to design a drug, one needs to design a lead compound binding to a specific target receptor rather than others to avoid side effects.[12,13] According to the definition of conventional specificity, one needs to compare all the competitive receptors in the human body binding with the same ligand. This makes the practical drug design for specificity impossible since it is impractical *in vitro* to evaluate binding affinity of one ligand binding to all possible competing targets. The structures of potential competing targets are not all readily available for the evaluation of the affinity, and even with the known structures, not all the affinities can be measured easily.

We proposed a novel concept to circumvent the challenge[14–16] (Fig. 1B). If the N- and C-terminus of multiple protein receptors are linked together with one another, then the resulting multiple connected receptors are equivalent to a single large receptor; *i.e.* the definition of the conventional thermodynamic specificity is transformed to the discrimination that a ligand binds to a specific site more favorably than other

[a]*Department of Chemistry & Physics, State University of New York at Stony Brook, Stony Brook, NY, 11790, USA. E-mail: jin.wang.1@stonybrook.edu; Fax: +1-631-632-7960; Tel: +1-631-632-1185*

[b]*State Key Laboratory of Electroanalytical Chemistry, Changchun Institute of Applied Chemistry, Chinese Academy of Sciences, Changchun, Jilin, 130022, China*

[c]*Graduate School of the Chinese Academy of Sciences, Beijing, 100039, China*

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c3sc50478f

‡ Zhiqiang Yan and Xiliang Zheng contributed equally to this work.

Published on 04 April 2013. Downloaded by Pennsylvania State University on 17/09/2016 00:16:25.

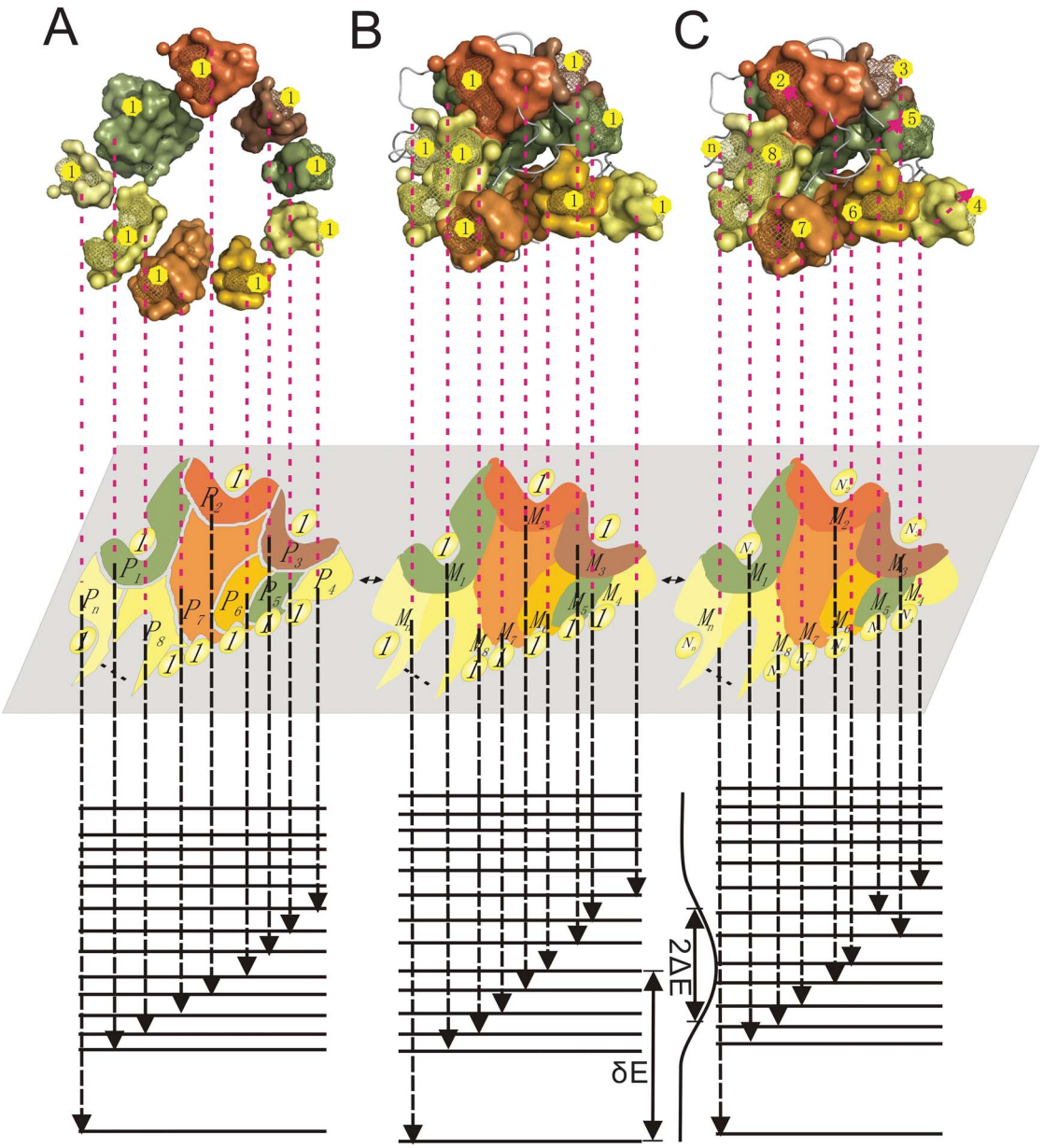

Fig. 1 Illustration of the equivalence of conventional and intrinsic specificities with three-dimensional view (top panel), two-dimensional view (medium panel) and corresponding energy spectrum (bottom panel). (A) A ligand (yellow circles) binding to multiple different receptors (P–P_n marked with different colors) showing the conventional specificity as the gap in binding affinity of the ligand binding to the specific receptor (P_n) in discrimination against other receptors. (B) Multiple different binding modes (M_1–M_n) of a ligand (yellow circles) on a large receptor showing the intrinsic specificity as the gap in binding affinity of the "native" binding mode (M_n) in discrimination against other binding modes. The large receptor can be thought as the multiple different receptors linked together. (C) Multiple different ligands (N_1–N_n) binding to the different or the same sites on the large receptor. The specificity here is the gap in binding affinity of a specific ligand (N_n) binding to the large receptor in discrimination against other ligands.

binding sites of this large receptor (Fig. 1A and B). Therefore the thermodynamic specificity becomes the existence of a gap in the binding free energy between the "native" and the rest for the ligand binding with different sites (binding modes) of the receptor or in other words the discrimination of "native" binding state (mode) from the rest. For the same reason, one receptor protein with different sites binding with a ligand is similar to the case of the whole universe of different ligands binding with the same receptor (Fig. 1B and C). If the target receptor protein is large enough, then this analogy should be exact. In other words, if the receptor protein is large enough, then probing the interactions of binding through different parts of the same target protein (Fig. 1B) and probing the interactions through the sequences (different receptors in Fig. 1A or different ligands in Fig. 1C) are equivalent. Under the assumption of large target protein (large enough to effectively represent the sequences of the receptor universe), we can term the new criterion for the specificity as the intrinsic specificity compared to the conventional definition of thermodynamic specificity. Thus, the thermodynamic specificity issue becomes

This journal is © The Royal Society of Chemistry 2013

Published on 04 April 2013. Downloaded by Pennsylvania State University on 17/09/2016 00:16:25.

the determination of intrinsic specificity of a ligand binding onto its receptor (Fig. 1B).

The process of protein–ligand binding can be physically quantified and visualized as a funnel-like energy landscape towards the native binding state with local roughness along the binding paths.[14,15,17–24] The "native" conformation of the complex is the conformation with the lowest binding energy. The energies of the decoy conformations follow a statistical Gaussian-like distribution (Fig. 1B). According to the theory of energy landscape for ligand binding,[14–16,25,26] the intrinsic specificity ratio (ISR = $\delta E/\Delta E\sqrt{S}$), where δE is the energy gap between the energy of "native" conformation and the average energy of decoy ensemble, ΔE is the energy fluctuation or the width of the energy distribution of the decoys, and S is the conformational entropy) is defined to quantify the magnitude of intrinsic specificity. The ISR indicates the capability of discriminating the "native" binding mode from other "non-native" modes for a particular ligand on its receptor, which can be readily quantified. Therefore, ISR physically provides a quantitative measure of the thermodynamic specificity without evaluating the conventional specificity through exploring the whole set of receptor universe.

Besides the significance of the thermodynamic specificity for ligand binding, there is an issue of kinetic specificity. A ligand can bind to multiple different receptors with very similar affinities but with different kinetics of on and/or off rates.[27,28] Essentially, the effectiveness and the duration of the biological activities are not only related to the thermodynamic stability and specificity, but also determined by the kinetic residence time that the ligand resides in the receptor. The longer the ligand is in residence at its receptor, the longer the biological effect endures *in vivo*. For the drug–target system, the residence time of a drug with its target is increasingly recognized to play a pivotal role in determining both the efficacy and toxicity of a drug and being as a critical predictor for drug efficacy and safety.[29–32] The residence time can discriminate different receptors even when the binding affinities to the same ligand are very close. Therefore the residence time can be used as a quantitative measure of the kinetic specificity and the selection criteria of the drug activity *in vitro*. However, the kinetic specificity is hard to be incorporated into drug screening *in silico* for lead compounds discovery. This is due to the fact that the computation of the residence time of a ligand on its receptor through molecular dynamics is very time-consuming, especially for the screening of a large library of ligands.[33,34] Here we describe the binding/unbinding kinetics as the diffusion process on the statistical binding energy landscape; the residence time as kinetic specificity can be directly computed and quantified by the kinetic time through diffusion (see Methods).

In this work, in order to accurately compute ISR as quantified intrinsic thermodynamic specificity and the residence time as quantified kinetic specificity, we employed our scoring function SPA (SPecificity and Affinity) which was developed recently by simultaneously optimizing the specificity and affinity.[16] SPA outperformed the current available scoring functions. By studying the ligands binding to the drug target model system of cyclooxygenases (COX-2 and COX-1), we found the statistical correlation between the intrinsic specificity and the conventional specificity. This validates the equivalence of intrinsic thermodynamic specificity and conventional thermodynamic specificity with a reasonable size receptor target protein. Furthermore, we found that the residence time as quantified kinetic specificity is correlated with the ISR of the thermodynamic specificity, and the binding affinity. This implies that thermodynamics and kinetics to some extent are cooperatively optimized in nature to ensure efficient and specific ligand binding, as well as the biological activity. With the quantified thermodynamic and kinetic specificities as selection criteria in addition to affinity for drug screening and target identification, we suggest a way of multidimensional (two- and three-dimensional) drug screening and target identification with affinity, thermodynamic intrinsic specificity and/or kinetic specificity. Our new strategy has important implications for drug discovery and design.

2 Results and discussion

2.1 Reliability of SPA scoring

In the structure-based drug discovery, scoring function describing protein–ligand interactions is responsible for the goodness of ranking in virtual screening. Docking programs as well as scoring functions have been developed over the past few decades for studying ligand binding. Successes in structural prediction and screening applications have been reported in some systems. However, accurate prediction of the binding affinities still remains challenging.[35–39] Furthermore, the specificity was never quantitatively addressed and implemented in the scoring function. To achieve a relatively reliable prediction of binding affinity and specificity, in this work, we employed our novel scoring function SPA in which affinity and intrinsic specificity were both taken into account for optimization[16] (see ESI†). In terms of the performance of testing the benchmark set, SPA outperformed other 16 popular scoring functions in academia and industry on the capability of identifying "native" binding conformation from "non-native" ones, as well as of predicting the binding affinities.[16] It demonstrates that the incorporation of the specificity to the strategy of developing scoring function greatly improves the performance over those given by the existing scoring functions that focused solely on affinity for development. SPA appears to be the best choice so far to calculate the energetics of protein–ligand interactions for affinity and the discrimination for specificity. To explore the reliability of SPA predictions on both affinity and specificity of COXs with small molecules, we performed docking studies of estimating binding affinities for a set of 20 drugs available on the market for the target COX-2 with known experimental affinities[40] (Table S1 of the ESI†). SPA gives a reasonable correlation coefficient of 0.71 between experimentally observed and predicted affinities (Fig. 2). Therefore, SPA can be reliably applied for further study of ligand binding and affinity prediction for COXs.

2.2 Connection of intrinsic and conventional specificities

As postulated in Fig. 1A and B, exploring binding interactions in the different binding sites of a receptor is equivalent to

Published on 04 April 2013. Downloaded by Pennsylvania State University on 17/09/2016 00:16:25.

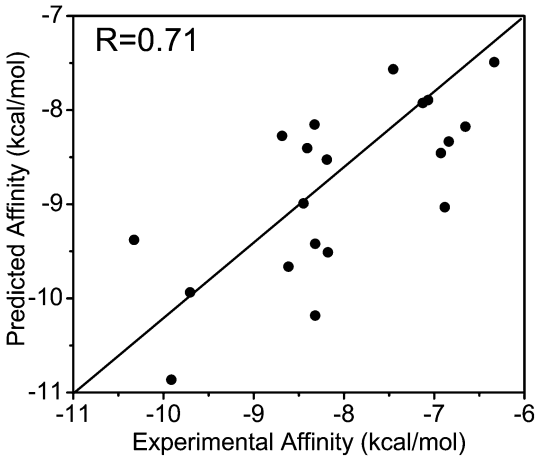

Fig. 2 Validation of SPA scoring. The statistical correlation (statistical significance $P < 0.001$) between experimental and predicted affinities of 20 drugs binding to COX-2 suggests the reliability and accuracy of SPA scoring for protein–ligand interactions.

exploring the interactions through the different receptors, if the target receptor is large enough to be thought as many smaller receptors linked together. Here, we tested this assumption by taking COX receptors as our example for docking, with 600 ligands from National Cancer Institute (NCI) diversity set. The correlation is shown (Fig. 3) between the conventional specificity (the affinity difference between COX-2 and its iso-form COX-1 with the same ligand) and intrinsic specificity ratio (ISR) for COX-2. The result indicates that the conventional specificity and intrinsic specificity quantified by ISR of COX-2 have significant correlations with each other. This validates the equivalence of conventional specificity and intrinsic specificity of a given ligand binding with a reasonable size receptor target (Fig. 1A and B). It also demonstrates that the readily available

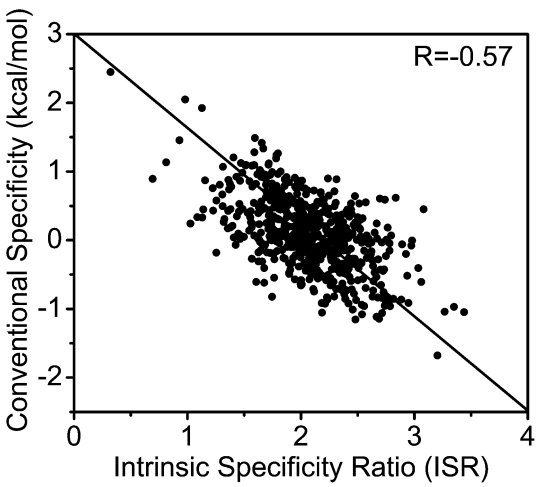

Fig. 3 Connection of intrinsic and conventional specificities. The statistical relationship between the conventional specificity (affinity difference between COX-2 and COX-1, $\Delta\Delta G = \Delta G_2 - \Delta G_1$) and intrinsic specificity ratio (ISR) of COX-2 binding with 600 ligands, validates the equivalence of conventional and intrinsic specificity.

ISR can be applied to explore the thermodynamic specificity of ligand binding. The non-perfect correlation may result from the fact that the number of target receptors are limited, with only COX-2 and COX-1. This is far fewer than the whole receptor universe, leaving the definition of the conventional specificity used here approximate and incomplete. The COX-2 protein is not large enough for the exact equivalence of the intrinsic specificity and the conventional specificity. A recent work on molecular dynamics of a ligand to search binding sites on the receptor also validates this assumption.[33] In that work, Shaw *et al.* carried out the longest explicit molecular dynamics simulations of protein–ligand binding so far, where the inhibitor PP1 was initially placed at a random location to search the binding sites on the protein Src kinase. Persistent and noteworthy intermediate conformations that the inhibitor locates diversely on the surface of the kinase were observed in the binding process, indicating competing binding modes do exist on the same receptor (Fig. 1B).

2.3 Connection of kinetics and thermodynamics

Essentially, the effectiveness and duration of biological activities are not only related to the thermodynamic stability and specificity, but also more directly determined by the kinetic residence time that the ligand resides on the receptor (off time).[27,28] The kinetic residence time can measure the duration of the drug on the target and directly determine the drug efficacy. With the same affinity, the kinetic residence time can be different and therefore provides another indicator of discriminating the binding of different receptors or different ligands. It is interesting to see how the kinetic specificity of the biomolecular recognition relates to the thermodynamic specificity quantified by ISR of the underlying binding energy landscape, as well as the thermodynamic stability. This is important for the field of drug discovery and design. First, residence time as kinetic specificity is increasingly characterized as the predictor of target selectivity and drug activity in addition to the affinity for thermodynamic stability.[31,32] Second, the theory of binding energy landscape quantifying the ISR provides a possible algorithm for designing ligands that can efficiently bind to the receptor.[14–16]

By describing the binding/unbinding kinetics as the diffusion process on the statistical energy landscape, we computationally quantified the residence time of a ligand on the receptor target as the kinetic specificity. The predicted kinetic specificity shows consistence with the experimental determined kinetic specificity for the COX-2 drugs with available residence times (Fig. 4A). This demonstrates the accuracy of our computation of residence time. A statistical correlation between the kinetic specificity (logarithm of residence time) and intrinsic thermodynamic specificity ISR is observed with the correlation coefficient 0.66 (Fig. 4B). This shows that the protein–ligand complex with a high ISR generally has a slow kinetics of unbinding. The correlation between kinetic specificity and ISR implies that the intrinsic thermodynamic specificity can be used as an indicator of residence time. The kinetic specificity and thermodynamic specificity are of different significance for

 This journal is © The Royal Society of Chemistry 2013

Published on 04 April 2013. Downloaded by Pennsylvania State University on 17/09/2016 00:16:25.

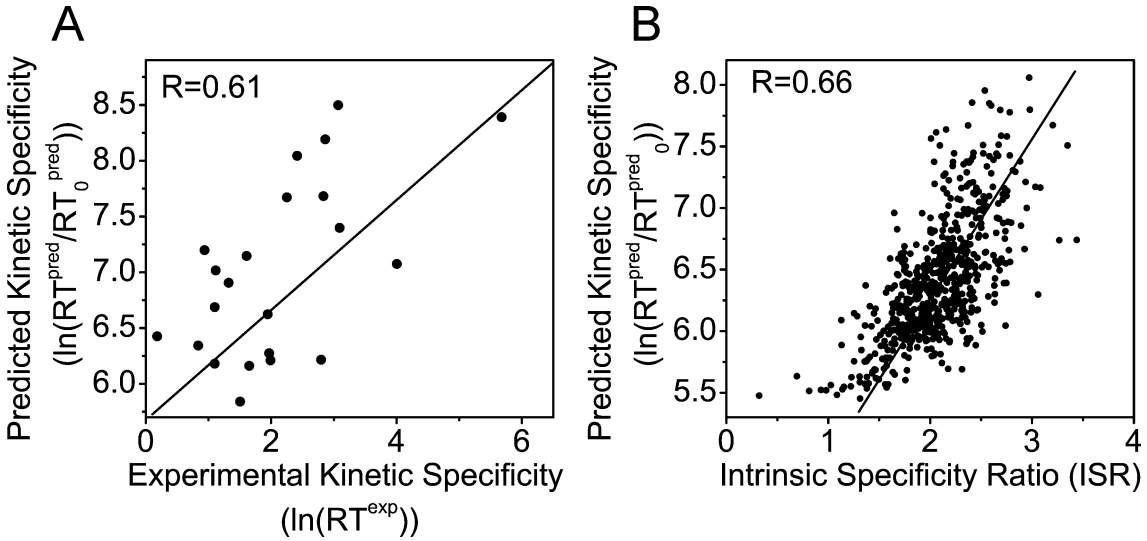

Fig. 4 Consistence of predicted and experimental kinetic specificities, and connection of the kinetic and thermodynamic specificities. (A) The predicted kinetic specificity (logarithm of RT^{pred}/RT_0^{pred}, RT^{pred} is the relative predicted residence time and RT_0^{pred} is the constant weighting factor) has relative high correlation (statistical significance $P < 0.01$) with the experimentally determined kinetic specificity (logarithm of RT^{exp}, RT^{exp} is the experimental residence time) for 22 drugs with available residence times. (B) The statistical relationship between the predicted thermodynamic specificity and kinetic specificity for 600 ligands of diversity set binding to COX-2, large ISR generally corresponds to long residence time.

biomolecular recognition, yet are correlated, suggesting the kinetics and thermodynamics of biomolecular recognition can be simultaneously optimized in nature for biological activities.

In practice, the thermodynamic intrinsic specificity ratio ISR is relatively easy to compute but difficult to experimentally measure. The kinetic specificity is relatively easy to experimentally measure but difficult to computationally quantify. Both the thermodynamic and kinetic specificities are related to the efficacy of drugs since the specificities of how well a drug discriminates the other compounds to a target in affinity and how long the drug interact with the target are major concerns for the effectiveness of the drug. The connection of residence time, as quantified kinetic specificity, and ISR, as quantified thermodynamic specificity, bridges the understanding and predicting of thermodynamics and kinetics of drug efficacy, and will undoubtedly be helpful for the field of drug discovery and design.

The thermodynamic view of funneled binding energy landscape[14,15] also implies that the specific protein–ligand complexes are at its global free energy minima relative to all other states. The binding affinity (or free energy) *in vitro* between a receptor and ligand is quantitatively assessed by the equilibrium measures of the ratio of the kinetic rate constants of complex association and dissociation. In equilibrium, there is an intimate link between the thermodynamic equilibrium constants and the kinetics. *In vivo*, however, the condition is often not in equilibrium and the effectiveness or the biological activity may no longer be determined by the affinity but by the residence time which can only be inferred by the kinetic measurement of rate of dissociation (off-rate constant).[28] It is important to know the correlation between the residence time and binding affinity. This can bridge the gap between *in vitro* measurements and *in vivo* biological efficacy. We observed that the predicted kinetic specificity quantified by the residence time is correlated to the affinity with certain statistical spread or dispersion (Fig. 5A), which is consistent with the correlation observed from experimental measurements for the COX-2 drugs with available residence times and affinities[41] (Fig. 5B and Table S1 of the ESI†). The correlation implies that the dissociation kinetics can be modulated by the specific interactions that influence the affinity.

2.4 Drug screening and target identification with thermodynamic and kinetic specificities

The funneled binding energy landscape[14,15] leads to the cooperative optimizations of thermodynamics and kinetics. However, we observed that the kinetic specificity has significant dispersion or spread at the same intrinsic specificity ratio (Fig. 4) or at the same affinity (Fig. 5A), and *vice versa*. In drug–target complex formation,[32] the residence time of a drug on its target results from the difference in binding free energy between the "native" and transition states. Efforts to improve the binding affinity of "native" states may reduce *in vivo* drug efficacy if stabilization of the "native" state is offset by the correspondingly larger stabilization of the transition state. Also, the improvement of binding stability could weaken the thermodynamic specificity if the energy gap between the specific and non-specific modes is decreased. This reinforces the importance of incorporating the intrinsic thermodynamic specificity ISR and kinetic specificity as additional probes for the biological activities.

Most often, whether experimentally or computationally, efforts for the identification of new lead compounds in drug discovery focused solely on the optimization of receptor–ligand affinity. This neglects the selection of kinetic and

Published on 04 April 2013. Downloaded by Pennsylvania State University on 17/09/2016 00:16:25.

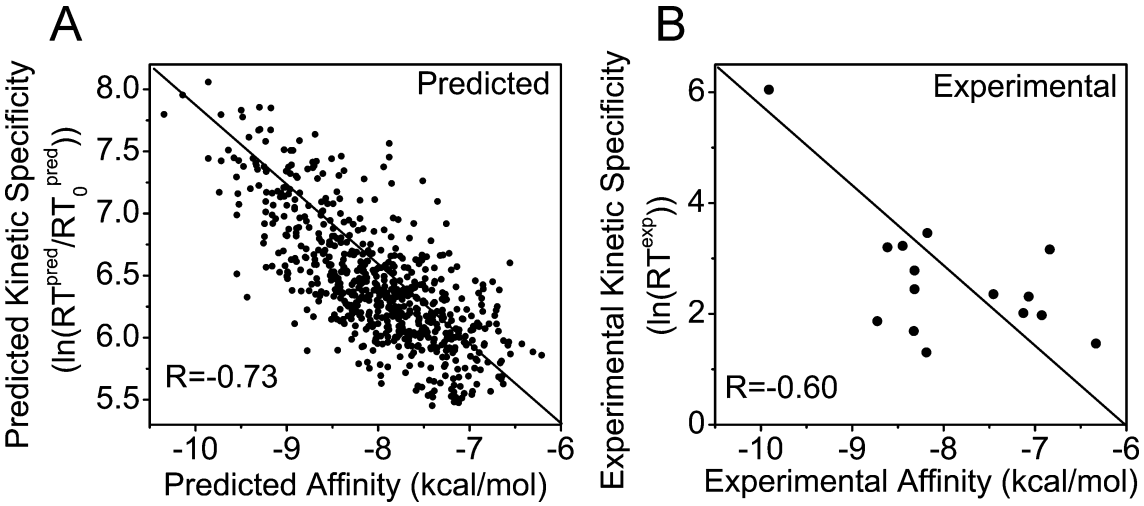

Fig. 5 Connection of the kinetic specificity and binding affinity. (A) The statistical relationship of kinetic specificities and affinities predicted by SPA for COX-2 binding with 600 ligands. (B) The statistical relationship (statistical significance $P < 0.05$) of kinetic specificities and affinities from available experimental measurements of 15 COX-2 drugs[40,41] (Table S1 of ESI†).

thermodynamic specificities we discussed above. We have seen (Fig. 6) that the compounds with both higher affinity and larger ISR are more likely to reside on the receptor with longer residence time. Likewise, the compounds with both higher affinity and longer residence time have more discrimination ability with a larger ISR. Whereas, residence time or ISR can be different at the same affinity, and *vice versa*. Thus, it is necessary to carry out multidimensional (two- or three-dimensional) drug screening and target identification by incorporating the kinetic and thermodynamic specificities as selection criteria, rather than affinity alone.

We chose two COX receptors as the model targets for drug screening and target identification test. Two COX isoenzymes (COX-2 and COX-1) are both responsible for the inflammation and pain.[42] Traditional nonsteroidal anti-inflammatory drugs (NSAIDs), are considered nonselective because they inhibit both COX-1 and COX-2. The inhibition of COX-2 by traditional NSAIDs accounts for the anti-inflammatory effect of the drugs while the inhibition of COX-1 can lead to NSAID toxicity and associated side effects, such as peptic ulceration. Thus inhibition of COX-1 is sometimes undesirable while inhibition of COX-2 is considered desirable. COX-2 selective drugs are a new form of NSAIDs that blocks just COX-2, leaving COX-1 to perform its essential jobs. These drugs are selective pain-killers and fever reducers, without the unpleasant side-effects.[43]

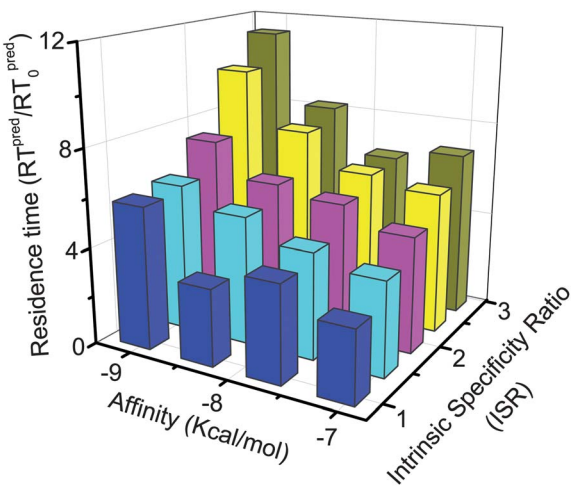

Fig. 6 Histogram of the residence time as a function of ISR and affinity. The values of these three parameters are extracted from the docking of 600 ligands onto the COX-2. Residence time or ISR can be different at the same affinity, and *vice versa*, suggesting the importance of incorporating both thermodynamics and kinetics into the drug screening and target identification.

To show the importance of incorporating thermodynamic and kinetic specificities into the drug screening and target identification, the discriminations of the selective drugs binding to the specific target COX-2 from its isoform COX-1, as well as the discriminations of the selective drugs from the non-selective drugs when binding to the COX-2, is quantified by statistical method Kolmogorov–Smirnov test (KS test, ESI†). The discrimination performance are compared among the single parameters (affinity (E), ISR or residence time (RT)) and the combined parameters through logistic regression (Fig. 7). The KS statistics (Fig. 7A) clearly shows that the single parameters can characterize the discriminations between the binding of selective drugs to specific target COX-2 from non-specific COX-1, with significant differences over 60%. It validates that thermodynamic specificity and kinetic specificity are important features of drug–target system, and are appropriate predictors for drug screening and even better than affinity. Furthermore, multidimensional screenings with the combinations of affinity and specificity (E + ISR, E + RT, or E + ISR + RT) exhibit more potential than any single parameter to discriminate the binding of COX-2 with selective drugs from COX-1. It is worth mentioning that three-dimensional screening is superior to two-dimensional screening to discriminate COX-2 from COX-1. The result of discrimination performance suggests that incorporation of thermodynamic and kinetic specificities can

 This journal is © The Royal Society of Chemistry 2013

Published on 04 April 2013. Downloaded by Pennsylvania State University on 17/09/2016 00:16:25.

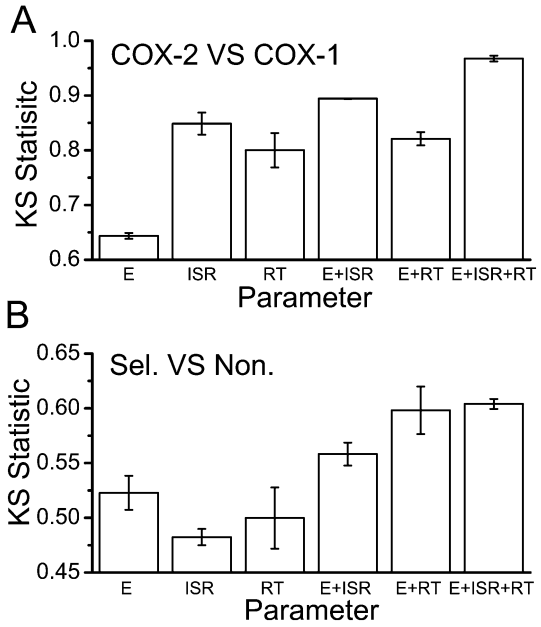

Fig. 7 Drug and target screening with thermodynamic and kinetic specificities. (A) The KS statistic of discrimination between the COX-2 and COX-1 when complexation with the selective drugs of COX-2. (B) The KS statistic of discrimination between the selective drugs and non-selective drugs upon complexation with COX-2.

effectively increase the discrimination between target receptor and other competitive receptors for a drug.

Also, significant differences between the selective drugs and non-selective drugs binding to COX-2 are also identified with the parameters (Fig. 7B). Collectively, multidimensional screening with a combination of affinity and specificity has higher capability than single parameter to discriminate the selective drugs from non-selective drugs. The enhanced discrimination performance by multidimensional screening is helpful to identify the selective drugs and increase the hit rate of lead compounds from compound library.

3 Conclusions

We have computationally quantified the thermodynamic and kinetic specificities of ligand binding. Experimentally, it is impractical to evaluate binding affinity of one ligand binding to all possible competing targets, and the structures of potential competing targets are not all readily available for evaluation of affinity. This limits the consideration of the conventional definition of the specificity for drug screening. ISR as quantified thermodynamic specificity is feasible to compute with fast docking and scoring methods *in silico* by generating the ensemble of docking decoys. It provides a practical way to select and optimize the leads by considering thermodynamic specificity in addition to affinity. The kinetic specificity is relatively easy to experimentally measure but difficult to computationally quantify. The residence time as quantified kinetic specificity is computed in our work with the kinetic unbinding time through diffusion on the statistical binding energy landscape. Based on the relationship between kinetics and thermodynamics quantified in our calculations, the computational two-dimensional virtual drug screening of ISR and affinity can greatly help to select drugs with longer residence time which is critical for the drug potency. Furthermore, one can combine the affinity, thermodynamic specificity and kinetic specificity for three-dimensional screening to increase the hit rate of discovering the optimized lead compounds and searching the receptor target of interest.

Our proposed concepts and methods on specificity have been tested on the drug–target system of COXs. COX-2 and COX-1 together provide an appropriate model system for validating our concepts and methods. The COX-2 and COX-1 are isoenzymes and competitive in binding to the small molecules. They have common non-selective drugs, and also uncommon selective drugs on the market. This makes COXs ideal for testing the equivalence of conventional specificity and intrinsic specificity, as well as for the virtual screening test. More importantly, the experimental data of binding affinities and residence times for some drugs are available to be compared with the predictions by our concepts and methods. The consistences with the available experimental measurements validate the accuracy of our predictions. Further, for the drug–target system of COXs, the remarkable performance on drug screening and target identification with quantified thermodynamic and kinetic specificities as screening criteria in addition to affinity is encouraging. This selection strategy can overcome the limitations of the screening through only the binding affinity. Our proposed concepts and methods, such as two- or three-dimensional virtual screening, provide a novel way to search the potential lead compounds. It is promising to apply our concepts and methods to other drug–target systems seeking for the inhibitors.

4 Materials and methods

4.1 Preparation of model system

The enzyme cyclooxygenase-2 (COX-2) was chosen as the model receptor protein which is the target of nonsteroidal antiinflammatory drugs (NSAIDs) for reducing fever and inflammation, such as the most taken drug aspirin. 600 small molecules were selected from the National Cancer Institute (NCI) Diversity Set database having molecular weights similar to that of the reference compound SC-558, with which the crystal structure of the COX-2 complex is available (PDB code 1CX2).[44,45] 30 COX-2 selective and 20 non-selective drugs (Table S1 of the ESI†) of NSAIDs were taken for the test of drug screening and target identification. COX-2 selective drugs are specific to inhibit only COX-2, while COX-2 non-selective drugs inhibit both the COX-2 and its isoenzyme COX-1 (PDB code 1Q4G).

4.2 Molecular docking and SPA scoring

Each ligand was docked to COX-2 and COX-1 by molecular docking with software AutoDock4.2.[46] Given that enough sampling of binding decoys is needed to generate the binding energy landscape, the grid box for ligand docking should be

Published on 04 April 2013. Downloaded by Pennsylvania State University on 17/09/2016 00:16:25.

sufficiently large to cover the active site as well as significant portions of the surrounding surface. The box size of docking was set as 100 × 100 × 100 grids centered at the active site with a grid spacing of 0.375. The box size is large enough to cover the active site as well as the surrounding surface. Within the grid box, Autodock4.2 stochastically generates a population of conformational, rotational, and translational isomers from the starting structure of the ligand and docks them with the conformational search method of Lamarckian genetic algorithm. During the search, the ligand was considered conformationally flexible with its torsional bonds defined by AutoDock4.2 according to their chemical features. Other parameters were set as the default values of AutoDock4.2. An ensemble of 2000 decoy conformations for each ligand were sampled. The energetics of decoys were scored by our developed scoring function SPA whose development strategy takes into account both specificity and affinity of protein–ligand complexes.[16] By clustering the decoys as the method implemented in AutoDock software in terms of their SPA scores and structural features, the decoy with the lowest energy in the largest population cluster was considered as the "native" conformation for the complex.

4.3 Computation of intrinsic specificity ratio

The intrinsic specificity ratio (ISR) for a given ligand with its receptor is calculated as

$$\mathrm{ISR} = \alpha\frac{\delta E}{\Delta E} \tag{1}$$

where α is a scaling factor which accounts for the contribution of the entropy to the specificity.[14] Here, it approximately depends on the number of torsional bonds of the ligands $\alpha \sim \sqrt{1/n_{\mathrm{tb}}}$. δE is the energy gap between the energy of native conformation E_{N} and the average energy of ensemble of decoys $\langle E_{\mathrm{D}}\rangle$, and ΔE is the energy fluctuation or the width of the energy distribution of the decoys, namely

$$\delta E = |E_{\mathrm{N}} - \langle E_{\mathrm{D}}\rangle| \tag{2}$$

$$\Delta E = \sqrt{\langle {E_{\mathrm{D}}}^2\rangle - \langle E_{\mathrm{D}}\rangle^2} \tag{3}$$

$\langle\ \rangle$ means the average over the ensemble of decoys.

4.4 Computation of residence time

To study the kinetics of ligand receptor binding, an order parameter is needed to describe the progress of binding/unbinding on the statistical energy landscape. In this work, the RMSD was employed as the order parameter (RMSD is the root mean square deviation relative to the "native" binding structure that represents the similarity to "native" state on the statistical energy landscape). Thermodynamically the energy landscape is characterized by the free energy which acts as a driving force for the binding process.[47–50] The kinetics of binding/unbinding on the thermodynamic energy landscape can be described by diffusion equation

$$\frac{\partial}{\partial t}P(r,t) = \frac{\partial}{\partial r}\left\{D(r)\left[\frac{\partial P(r,t)}{\partial r} + P(r,t)\frac{\partial(F(r)/k_{\mathrm{B}}T)}{\partial r}\right]\right\} \tag{4}$$

where $P(r, t)$ is the probability of the biomolecular binding complex with specific r (RMSD) at time t, $D(r)$ is the diffusion coefficient projected to r, and $F(r)$ is the free energy of the system at a given r. In terms of the order parameter RMSD, the problem becomes one-dimensional diffusion. The diffusion equation can be integrated to give unbinding time or residence time (RT) for the "native" binding mode by:

$$\mathrm{RT} = \int_{r}^{r_{\mathrm{u}}} \mathrm{d}r \int_{r_{\mathrm{n}}}^{r'} \mathrm{d}r' \frac{\exp\left[\dfrac{F(r) - F(r')}{k_{\mathrm{B}}T}\right]}{D(r)} \tag{5}$$

The boundary conditions are set as a reflecting one at $r_{\mathrm{n}} = 0$, where system is in "native" state, and an absorbing one at r_{u}, where the system is in the "non-native" states. The choice of r_{u} depends on the system. For simplicity, $D^{-1}(r)$ was set as constant and the order parameter RMSD was normalized and discretized into a set of bins. By using the discrete form of the full double integral, the residence time is calculated as:

$$\mathrm{RT} = \frac{1}{D}\sum_{m}^{M}\sum_{0}^{m'}\exp\left[\frac{F(m) - F(m')}{k_{\mathrm{B}}T}\right] \tag{6}$$

where M is the total number of bins and set as 20.

Acknowledgements

JW thanks National Science Foundation for support. XLZ and EKW are supported by the National Natural Science Foundation of China (Grants 21190040, 11174105 and 91227114). The authors thank Jeremy Adler for editing language and grammar.

References

1 J. A. McCammon, *Curr. Opin. Struct. Biol.*, 1998, **8**, 245–249.
2 N. C. Cohen, *Guidebook on molecular modeling in drug design*, Academic Press, New York, 1996, pp. 1–361.
3 E. Robert and S. L. Bender, *Chem. Rev.*, 1997, **97**, 1359–1472.
4 G. Kar, O. Keskin, A. Gursoy and R. Nussinov, *Curr. Opin. Pharmacol.*, 2010, **10**, 715–722.
5 W. Tan, H. Wang, Y. Chen, X. Zhang, H. Zhu, C. Yang, R. Yang and C. Liu, *Trends Biotechnol.*, 2011, **29**, 634–640.
6 J. Janin, *Biochimie*, 1995, **77**, 497–505.
7 J. J. Havranek, P. B. Harbury, *et al.*, *Nat. Struct. Biol.*, 2002, **10**, 45–52.
8 T. Kortemme, L. A. Joachimiak, A. N. Bullock, A. D. Schuler, B. L. Stoddard and D. Baker, *Nat. Struct. Mol. Biol.*, 2004, **11**, 371–379.
9 D. N. Bolon, R. A. Grant, T. A. Baker and R. T. Sauer, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 12724–12729.
10 G. Grigoryan, A. W. Reinke and A. E. Keating, *Nature*, 2009, **458**, 859–864.
11 S. J. Fleishman and D. Baker, *Cell*, 2012, **149**, 262–273.
12 R. L. Chang, L. Xie, L. Xie, P. E. Bourne and B. Ø. Palsson, *PLoS Comput. Biol.*, 2010, **6**, e1000938.

 This journal is © The Royal Society of Chemistry 2013

Published on 04 April 2013. Downloaded by Pennsylvania State University on 17/09/2016 00:16:25.

13 E. Karakoc, A. Cherkasov and S. C. Sahinalp, *Bioinformatics*, 2006, **22**, e243–e251.
14 J. Wang and G. M. Verkhivker, *Phys. Rev. Lett.*, 2003, **90**, 188101.
15 J. Wang, X. Zheng, Y. Yang, D. Drueckhammer, W. Yang, G. Verkhivker and E. Wang, *Phys. Rev. Lett.*, 2007, **99**, 198101.
16 Z. Yan and J. Wang, *Sci. Rep.*, 2012, **2**, 309.
17 J. D. Bryngelson, J. N. Onuchic, N. D. Socci and P. G. Wolynes, *Proteins: Struct., Funct., Genet.*, 1995, **21**, 167–195.
18 J. Janin, *Proteins: Struct., Funct., Genet.*, 1996, **25**, 438–445.
19 P. A. Rejto and G. M. Verkhivker, *Proc. Natl. Acad. Sci. U. S. A.*, 1996, **93**, 8945–8950.
20 D. W. Miller and K. A. Dill, *Protein Sci.*, 2008, **6**, 2166–2179.
21 C.-J. Tsai, S. Kumar, B. Ma and R. Nussinov, *Protein Sci.*, 1999, **8**, 1181–1190.
22 B. N. Dominy and E. I. Shakhnovich, *J. Med. Chem.*, 2004, **47**, 4538–4558.
23 Z. Liu, B. N. Dominy and E. I. Shakhnovich, *J. Am. Chem. Soc.*, 2004, **126**, 8515–8528.
24 Y. Levy, P. G. Wolynes and J. N. Onuchic, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 511–516.
25 D. U. Ferreiro, J. A. Hegler, E. A. Komives and P. G. Wolynes, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 3499–3503.
26 D. U. Ferreiro, J. A. Hegler, E. A. Komives and P. G. Wolynes, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 19819–19824.
27 G. Vauquelin and I. Van Liefde, *Trends Pharmacol. Sci.*, 2006, **27**, 355.
28 P. J. Tummino and R. A. Copeland, *Biochemistry*, 2008, **47**, 5481–5492.
29 D. C. Swinney, *Nat. Rev. Drug Discovery*, 2004, **3**, 801–808.
30 D. C. Swinney, *Lett. Drug Des. Discovery*, 2006, **3**, 569–574.
31 R. A. Copeland, D. L. Pompliano and T. D. Meek, *Nat. Rev. Drug Discovery*, 2006, **5**, 730–739.
32 H. Lu and P. J. Tonge, *Curr. Opin. Chem. Biol.*, 2010, **14**, 467–474.
33 Y. Shan, E. T. Kim, M. P. Eastwood, R. O. Dror, M. A. Seeliger and D. E. Shaw, *J. Am. Chem. Soc.*, 2011, **133**, 9181–9183.
34 E. A. Proctor, S. Yin, A. Tropsha and N. V. Dokholyan, *Biophys. J.*, 2012, **102**, 144–151.
35 D. B. Kitchen, H. Decornez, J. R. Furr and J. Bajorath, *Nat. Rev. Drug Discovery*, 2004, **3**, 935–949.
36 V. Mohan, A. C. Gibbs, M. D. Cummings, E. P. Jaeger and R. L. DesJarlais, *Curr. Pharm. Des.*, 2005, **11**, 323–333.
37 K. Onodera, K. Satou and H. Hirota, *J. Chem. Inf. Model.*, 2007, **47**, 1609–1618.
38 I. W. Davis and D. Baker, *J. Mol. Biol.*, 2009, **385**, 381–392.
39 H. Jenssen, C. D. Fjell, A. Cherkasov and R. E. Hancock, *J. Pept. Sci.*, 2008, **14**, 110–114.
40 G. Dannhardt and W. Kiefer, *Eur. J. Med. Chem.*, 2001, **36**, 109–126.
41 C. Knox, V. Law, T. Jewison, P. Liu, S. Ly, A. Frolkis, A. Pon, K. Banco, C. Mak, V. Neveu, *et al.*, *Nucleic Acids Res.*, 2010, **39**, D1035–D1041.
42 W. L. Smith, D. L. DeWitt and R. M. Garavito, *Annu. Rev. Biochem.*, 2000, **69**, 145–182.
43 A. K. Chakraborti, S. K. Garg, R. Kumar, H. F. Motiwala and P. S. Jadhavar, *Curr. Med. Chem.*, 2010, **17**, 1563–1593.
44 R. G. Kurumbail, A. M. Stevens, J. K. Gierse, J. J. McDonald, R. A. Stegeman, J. Y. Pak, D. Gildehaus, J. M. Miyashiro, T. D. Penning, K. Seibert, *et al.*, *Nature*, 1996, **384**, 644–648.
45 D. L. DeWitt, *Mol. Pharmacol.*, 1999, **55**, 625–631.
46 D. S. Goodsell, G. M. Morris and A. J. Olson, *J. Mol. Recognit.*, 1996, **9**, 1–5.
47 J. Wang, W. Huang, H. Lu and E. Wang, *Biophys. J.*, 2004, **87**, 2187–2194.
48 J. Wang, *Chem. Phys. Lett.*, 2006, **418**, 544–548.
49 J. D. Bryngelson and P. G. Wolynes, *Proc. Natl. Acad. Sci. U. S. A.*, 1987, **84**, 7524–7528.
50 N. Socci, J. Onuchic and P. Wolynes, *J. Chem. Phys.*, 1996, **104**, 5860–5868.