Science Advances

AAAS

advances.sciencemag.org/cgi/content/full/2/3/e1501240/DC1

Supplementary Materials for

Regulation of protein-ligand binding affinity by hydrogen bond pairing

Deliang Chen, Numan Oezguen, Petri Urvil, Colin Ferguson, Sara M. Dann, Tor C. Savidge

Published 25 March 2016, *Sci. Adv.* **2**, e1501240 (2016)
DOI: 10.1126/sciadv.1501240

The PDF file includes:

Text S1. Theoretical proof of the H-bond pairing principle.
Text S2. Relationship between the free energy change for a reversible protein-ligand H-bond competing process and the H-bonding capability of the H-bond–forming atoms.
Text S3. The H-bonding capability of the protein atoms with which a ligand atom interacts and the effect of H-bond geometry on the H-bond interaction.
Fig. S1. Schematic illustration of the free energy change for the H-bond competing process.
Fig. S2. Calculation of H-bonding capability based on water/hexadecane partition coefficients.
Fig. S3. Contributions of the H-bonds between CN and the Tyr-OH from scytalone dehydratase to protein-ligand interactions.
Fig. S4. Proof for the strong H-bond interactions between the CN group of inhibitor **2** and tyrosine hydroxyls from scytalone dehydratase.
Fig. S5. Binding affinities of 1*H*-imidazole-2-sulfonamide and thiophene-2-sulfonamide.
Fig. S6. Quaternary ammonium cation [$–N(Me)_3^+$]-π interactions are more favorable than ammonium ion ($–NH_3^+$)-π interactions.
Fig. S7. A pathogenic role for the s-s/w-w H-bonding principle in melanin toxicity.
Fig. S8. The thermodynamic cycle that demonstrates the contribution of H-bonds to enzymatic catalytic power equates to their contribution to protein-ligand binding.
References (*46–52*)

Text S1. Theoretical Proof of the H-bond Pairing Principle

I: Proof in Enthalpy.

1: Overview of the proof: Assume two H-bond acceptors (A_1 and A_2) and two H-bond donors (D_1-H and D_2-H) are in a solution, two acceptors and two donors form mixed pairing. A general H-bond competing equation is shown in equation (S1)

$$D_1\text{-}H\ldots A_2 + D_2\text{-}H\ldots A_1 \leftrightarrow D_1\text{-}H\ldots A_1 + D_2\text{-}H\ldots A_2 \qquad \text{(S1)}$$

D_2-H and A_2 are assumed to have stronger H-bonding capabilities than D_1-H and A_1, respectively. The aim of this proof is to determine whether the free energy change (ΔG) of equation (S1) favors the side of D_2-H…A_2 using assumed relative strengths of H-bonds in equation (S1) at optimal H-bond distances.

2. Function for calculating the strength of H-bond D-H....A:

H-bond energy (E) is usually considered the sum of non-covalent van der Waals (E_{vdw}) and electrostatic interactions (E_c) (46-48). E_{vdw} is calculated using the equilibrium distance (r_{oij}) and well depth (ε_{ij}) as shown in the first summation of equation (S2), while E_c is calculated based on Coulomb's law (second sum).

$$E = \sum \varepsilon_{ij}\left[\left(\frac{r_{oij}}{r_{ij}}\right)^{12} - 2\left(\frac{r_{oij}}{r_{ij}}\right)^{6}\right] + \sum \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \qquad \text{(S2)}$$

As we are interested in the relative strengths of H-bonds, we make the following assumptions for simplification:

(i) For E_{vdw}, we consider only the van der Waals interaction between A and H, because it is much larger than the other van der Waals interactions (rAH<rDA).

(ii) E_c is directly proportional to the charge A(q_A) and is inversely proportional to the distance r_{AH} between H and A (*Appendix I*):

$$E_c \propto q_A \text{ and } E_c \propto 1/r_{AH}$$

For simplification, r represents the distance between A and H. Thus, equation (S2) can be expressed as:

$$E = \varepsilon\left(\frac{r_0^{12}}{r^{12}} - \frac{2*r_0^6}{r^6}\right) + E_c \qquad (E_c \propto q_A; E_c \propto 1/r) \qquad (S3)$$

(iii) The H-bond compete pairing process is in solution, it is reasonable to assume the H-bonds are at their optimal distances and it is not necessary to consider H-bond angles.

3. Strength of H-bond D-H....A at optimal distance (E_m):

H-bond energy is minimal at optimal H-bond distance. Therefore, the derivative of the energy is zero (d(E)/d(r) = 0). The derivative of Equation (S3 with respect to r gives the following equation of E_m (*Appendix II*):

$$E_m = \frac{11*E_c}{12} - \frac{\epsilon*r_0^6}{r^6} \qquad (S4)$$

4. Difference in E_m between D-H...A_1 and D-H...A_2:

Assume A_1 and A_2 have the same atom type and their charges are q_A and $k*q_A$ (k>1), respectively. Their E_m values and the difference $\triangle E$ [E_m(D-H ....A_2) - E_m(D-H ....A_1)] are calculated as follows:

$$D_1\text{-H...}A_1\ (r_1) \xrightarrow[\triangle E_1]{q_{A2} = k*q_{A1}} D_1\text{-H...}A_2\ (r_2)$$

E_c	E_c	$E_{2c}\{= E_c \times k \times \left(\frac{r_1}{r_2}\right)\}$
E_m	$\frac{11\times E_c}{12} - \frac{\varepsilon\times r_0^6}{r_1^6}$	$\frac{11\times E_{2c}}{12} - \frac{\varepsilon\times r_0^6}{r_2^6}$

$\triangle E_1$, the difference between E_m(D_1-H ....A_2) and E_m(D_1-H ....A_1), is:

$$\Delta E_1 = \frac{11}{12} \times E_c\left(\frac{k\times r_1}{r_2} - 1\right) - \varepsilon \times r_0^6 \times \frac{1}{r_1^6}\left(\frac{r_1^6}{r_2^6} - 1\right) \qquad (S5)$$

Similarly, $\triangle E_2$ for D_2-H ....A_1 and D_2-H....A_2 is:

$$\Delta E_2 = \frac{11}{12} \times E_c'\left(\frac{k\times R_1}{R_2} - 1\right) - \varepsilon \times r_0^6 \times \frac{1}{R_1^6}\left(\frac{R_1^6}{R_2^6} - 1\right) \qquad (S6)$$

where r_1, r_2, R_1 and R_2 are distances between H and A for the H-bonds D_1-H ....A_1, D_1-H ....A_2, D_2-H ....A_1, and D_2-H ....A_2, respectively. E_c and E_c' are the electrostatic interactions for D_1-H....A_1 and D_2-H....A_1. $E_c'$is stronger than E_c ($-E_c' > -E_c$ or $E_c' < E_c$), because D_2-H is assumed to have stronger H-bonding capability than D_1-H.

5. Relationship between ΔE_1 and ΔE_2:

As $-E_c' > -E_c$, we can prove that $R_1/R_2 > r_1/r_2 > 1$ (*Appendix III*). By comparing each term between ΔE_1 (equation (S5)) and ΔE_2 (equation (S6)), we can conclude that ΔE_2 is lower than ΔE_1 ($-\Delta E_2 > -\Delta E_1$or $\Delta E_2 - \Delta E_1 < 0$).

6. Enthalpy change for equation (S1):

$$\Delta H = E_m(H\text{-}D_2...A_2) + E_m(H\text{-}D_1...A_1) - E_m(H\text{-}D_2\ldots A_1) - E_m(H\text{-}D_1\ldots A_2)$$

$$= \Delta E_2 - \Delta E_1 < 0 \qquad \text{(S7)}$$

Therefore, the H-bond competing process favors the s-s/w-w pairing side in enthalpy.

7. Proof when A_1 and A_2 in equation (S1) are not the same atom type:

If A_1 and A_2 are not the same atom type, the equilibrium distances (r_{oij}) and well depths (ε_{ij}) are unequal, and equations (S5) and (S6) need to be modified. To prove whether the s-s/w-w H-bond pairing statement still holds, a C++ program was developed to calculate enthalpy changes of Equation (S1 for all possible combinations of O, N and S atoms (*Appendix IV)*. Results show that equation (S1) favors the side of the strongest H-bond.

II. Proof in Free Energy:

The H-bond interaction between H-bond donor and acceptor causes orientational and positional restrictions of the H-bond forming atoms. Thus, the entropy-enthalpy compensation always exist in the protein-ligand H-bonding. The calculation results (5) based on classical statistical mechanics for the TΔS and ΔH contributions to ΔG for different well depths indicate that some compensation exists between TΔS and ΔH. For H-bonds at its optimal distances, the favorable enthalpic (ΔH) contributions are only partially cancelled by unfavorable entropic (TΔS) contributions, which means $T\Delta S/\Delta H < 1$ (rule 1). The slope in the TΔS ~ ΔH plot for weaker H-bond is larger than that that for stronger H-bond (rule 2). The TΔS ~ ΔH plot is shown in Figure S1B, in which the H-bond energy at optimal distance (Em) represents ΔH.

A H-bond competing process and the enthalpy change of this process are shown in equation (S8):

$$sw1 + sw2 \rightarrow ss + ww \qquad \Delta H = Em_{ss} + Em_{ww} - Em_{sw2} - Em_{sw1} < 0 \qquad (S8)$$

s: denotes strong H-bond forming capability; w: denotes weak H-bond forming capability; Em: refers to the H-bond energy at optimal distances. Em_{sw1}, Em_{sw2}, Em_{ss}, and Em_{ww} represent E_m values for sw1, sw2, ss and ww, respectively. The Em values are all negative with the strongest H-bond having the lowest E_m value. To evaluate the effect of entropy change on the free energy change of the H-bond competing process, Equation (S8 is decomposed into two processes as shown in equations (S9) and (S10):

$$sw1 + sw2 \rightarrow ss^* + ww \qquad \Delta H^1 = Em_{ss^*} + Em_{ww} - Em_{sw2} - Em_{sw1} = 0 \qquad (S9)$$

$$ss^* \rightarrow ss \qquad \Delta H^2 = \Delta H \qquad (S10)$$

where ss* is an imaginary H-bond which is different from the ss in that the ss* H-bond distance is larger than the optimal distance and thus the ss* H-bond is weaker than the ss H-bond. The TΔS ~ Em plots for the H-bonds sw1, sw2, ss, ww and ss* are shown in Figure S1B.

For equation (S9):

Because $Em_{ss*} + Em_{ww} - Em_{sw2} - Em_{sw1} = 0$, thus: $Em_{ww} - Em_{sw1} = Em_{sw2} - Em_{ss*}$.

Because the slope in the $T\Delta S \sim \Delta H$ plot for weaker H-bond is larger than that that for stronger H-bond (rule 2), we can get:

$$\frac{T\Delta S_{ww} - T\Delta S_{sw1}}{Em_{ww} - Em_{sw1}} > \frac{T\Delta S_{sw2} - T\Delta S_{ss*}}{Em_{sw2} - Em_{ss*}}$$

Because $Em_{ww} - Em_{sw1} = Em_{sw2} - Em_{ss*} > 0$; we can get:

$$T\Delta S_{ww} - T\Delta S_{sw1} > T\Delta S_{sw2} - T\Delta S_{ss*}$$

Thus, the ΔG for Equation S9 is: $\Delta H^1 - T^*(\Delta S_{ww} + \Delta S_{ss*} - \Delta S_{sw1} - \Delta S_{sw2} - T) < 0$ (S11)

For equation (S10):

Because $T\Delta S/\Delta H < 1$ (rule 1), thus $\Delta H < T\Delta S$ and $\Delta H - T\Delta S < 0$;

Therefore, the ΔG for equation (S10) is less than 0.

As ΔG for equation (S8) is the sum of the ΔG for equation (S9) and the ΔG for equation (S10), ΔG for equation (S8) is less than 0. So, the H-bond competing process favors the s-s/w-w pairing H-bonds in free energy.

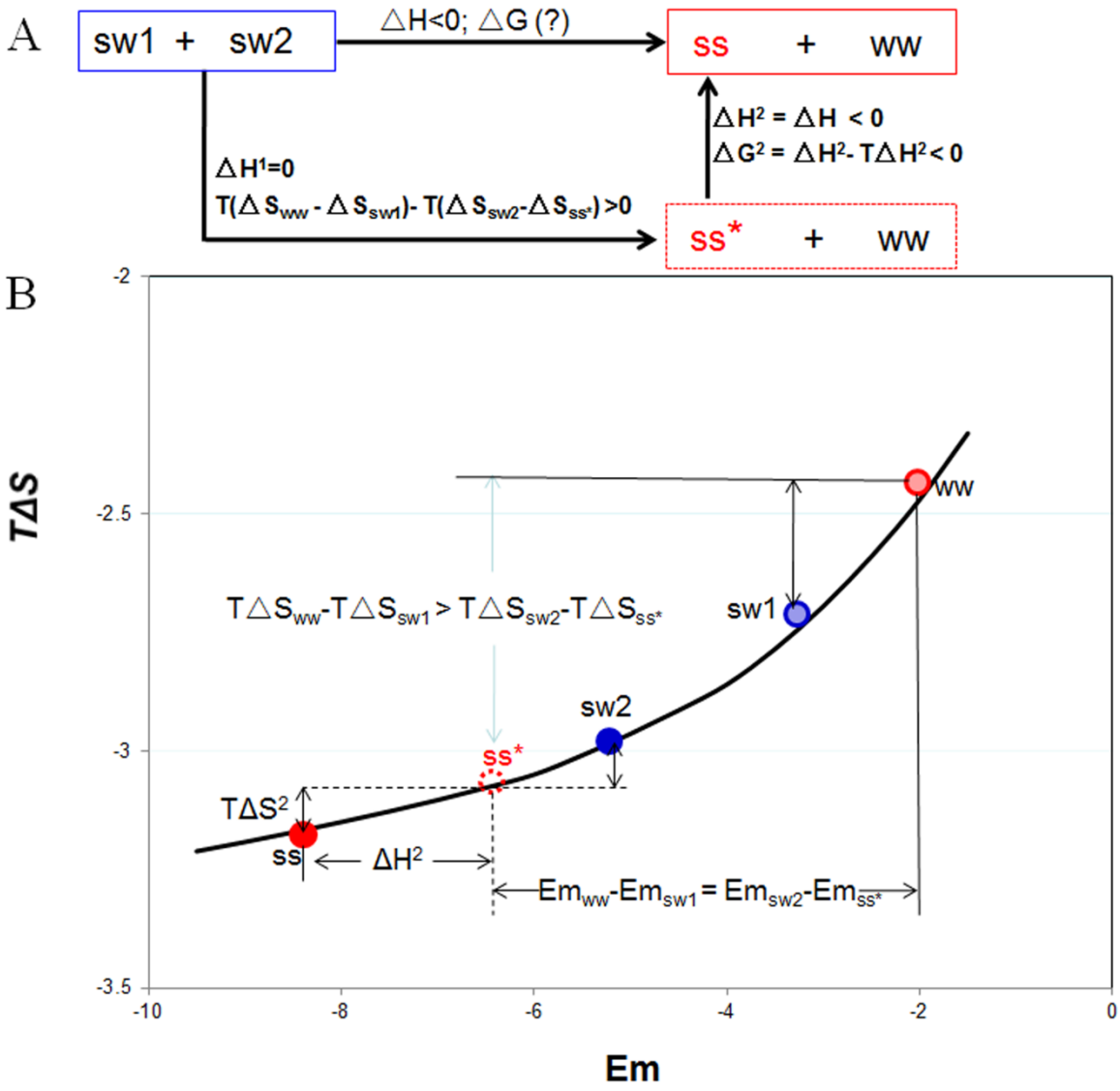


fig. S1. Schematic illustration of the free energy change for the H-bond competing process. (A) The H-bond competing process is composed of two processes and the ΔH of the first process is 0. (B) The relationship between TΔS required to form H-bonds and the E_m values of the H-bonds. The ΔH and TΔS for the processes in (A) can be estimated based on the relationship.

Proof Appendix:

Appendix I. To prove: $E_c \propto 1/r_{AH}$:

E_c includes all electrostatic interactions. For the H-bond D-H…A assume charges are q_D, q_H and q_A for D, H and A respectively; and distances for D-A and H-A are r_{DA} and r_{AH} respectively. Thus, the Ec for D-H…A is:

$$E_C = \frac{q_D \times q_A}{4\pi\epsilon_{DA} r_{DA}} + \frac{q_H \times q_A}{4\pi\epsilon_{HA} r_{HA}} = \frac{q_A}{4\pi\epsilon_{HA} r_{HA}} \left(q_D \times \frac{\epsilon_{HA}}{\epsilon_{DA}} \times \frac{r_{HA}}{r_{DA}} + q_H \right)$$

We can define an effective point charge q'_H at the position of H as:

$$q'_H = q_D \times \frac{\epsilon_{HA}}{\epsilon_{DA}} \times \frac{r_{HA}}{r_{DA}} + q_H$$

q'_H changes with H-bond distances r_{HA} and r_{DA}. However, this change is negligible. Consider r_{HA} changes from 2.0 (medium strength H-bond) to 1.5 (strong H-bond) and r_{DA} changes from 3.0 to 2.5:

$$\Delta q'_H = q_D \times \frac{\epsilon_{HA}}{\epsilon_{DA}} \left(\frac{2.0}{3.0} - \frac{1.5}{2.5} \right) = 0.067 \times q_D \times \frac{\epsilon_{HA}}{\epsilon_{DA}}$$

ε_{HA} is significantly less than ε_{DA} (otherwise, the water-water H-bond is repulsive). Therefore, $\Delta q'_H$ (which is considerably lower than q'_H, and q_H) can be regarded as constant and there is no need to calculate an accurate strength of the H-bond.

Thus, it is reasonable to assume that $E_c \propto 1/r_{AH}$.

Appendix II. To prove:

$$E_m = \frac{11 * E_c}{12} - \frac{\epsilon \times r_0^6}{r^6}$$

The derivative of equation (S3) with respect to r gives the following equation:

$$-12\,\varepsilon\,\frac{r_0^{12}}{r^{13}} + 12\,\varepsilon\,\frac{r_0^6}{r^7} - \frac{E_c}{r} = 0 \qquad \text{(b.1)}$$

$$\varepsilon\,\frac{r_0^{12}}{r^{12}} - \varepsilon\,\frac{r_0^6}{r^6} = -\frac{E_c}{12} \qquad \text{(b.2)}$$

$$E_{vdw} = \epsilon\left(\frac{r_0^{12}}{r^{12}} - \frac{2\times r_0^6}{r^6}\right) = -\frac{E_C}{12} - \frac{\varepsilon\times r_0^6}{r^6} \quad \text{(b.3)}$$

Therefore: $E_m = \frac{11*E}{12} - \frac{\epsilon\times r_o^6}{r^6}$ (b.4)

Appendix III. To prove: $R_1/R_2 > r_1/r_2 > 1$

Based on the equation (b.2) for H-bond D_1-H ....A_2 and for D_1-H....A_1, the following formula is obtained:

$$\frac{\frac{r_0^{12}}{r_2^{12}} - \frac{r_0^6}{r_2^6}}{\frac{r_0^{12}}{r_1^{12}} - \frac{r_0^6}{r_1^6}} = k\frac{r_1}{r_2} \quad \text{(c.1)}$$

Let: f = r1/r2: the above formula becomes:

$$\frac{r_0^6\times f^{11} - r_1^6\times f^5}{r_0^6 - r_1^6} = k \quad \text{(c.2)}$$

The following equation can be obtained from equation (c.2):

$$\frac{r_1^6}{r_0^6} = \frac{k - f^{11}}{k - f^5} = 1 - f^5 \times \frac{f^6 - 1}{k - f^5} \quad \text{(c.3)}$$

Let: $F = R_1/R_2$. Based on D_2-H ....A_1 and D_2-H ....A_2, a similar equation can be obtained:

$$\frac{R_1^6}{r_0^6} = 1 - F^5 \times \frac{F^6 - 1}{k - F^5} \quad \text{(c.4)}$$

Because $R_1 < r_1$, F must be greater than f and the right side of equation (c.4) is less than the right side of equation (c.3). As r1 > r2, r1/r2 > 1

Therefore, $R_1/R_2 > r_1/r_2 > 1$

Appendix IV. *Method for calculating the enthalpy changes of equation (S1) for all possible combinations of O, N and S atoms using the program C++*:

Algorithm:

1: Set values and parameters:

(A) q'_{H1} and q'_{H2} (charges of D_1-H and D_2-H): minimum: 0.02; maximum: 1; increment: 0.02;

q_{A1} and q_{A2} (charges of A_1 and A_2): minimum: -1; maximum: -0.02; increment: 0.02;

dielectric constant (D): 4 and 78.5;

(B) van der Waals parameters:

N and H: equilibrium distance: 2.75Å; well depth: 0.057kcal/mol;

O and H: equilibrium distance: 2.60Å; well depth: 0.063kcal/mol;

S and H: equilibrium distance: 3.00Å; well depth: 0.063kcal/mol;

2) Calculate E_m for H-bond D-H....A with q_H for D-H and q_A for A:

Based on Coulomb's law: $E_c = k*q_H*q_A /(D*r)$

(k: 332 kcal/mol, D: dielectric constant) and equation (b.2) (in Appendix II)

$$\varepsilon\frac{r_0^{12}}{r^{12}} - \varepsilon\frac{r_0^{6}}{r^{6}} = -\frac{E_C}{12} \qquad \text{(b.2, in Appendix II)}$$

The following equation can be obtained:

$$\frac{12\times\epsilon\times r_0^{12}}{r^{11}} - \frac{12\times\epsilon\times r_0^{6}}{r^{5}} = -k\times\frac{q_H\times q_A}{D}$$

The algorithm involves changing the value of r and calculating the value of the left side until the relative difference between the both sides is less than 0.0000001.

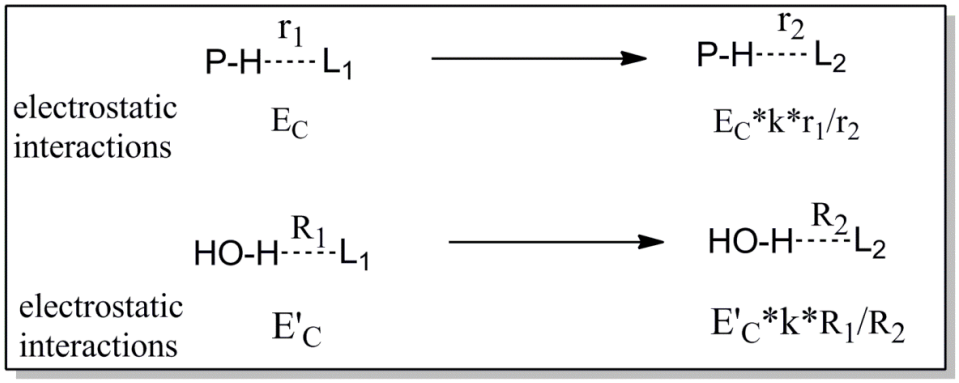
After r is obtained, E_m can be obtained.

3) Calculate ΔE for equation (S1), and the enthalpy change does not favor the side of the strongest H-bond, record the result.

Results: A total of 3.75 x 10^7 H-bond competing pairing processes were calculated. Only 8.36 x 10^4 (0.23%) H-bond competing processes did not favor the strongest H-bond in enthalpy. In these cases, E_m values for the strongest and second strongest H-bonds were almost identical (*e.g.* -11.4492 kcal mol^{-1} *vs.* -11.4488 kcal mol^{-1}). The largest ΔH for the processes measured was 0.0086 kcal mol^{-1}, less than 0.1% of the strongest H-bond (-11.4492 kcal mol^{-1}).

Text S2. Relationship between the free energy change (△G) for a reversible protein-ligand H-bond competing process and the H-bonding capability of the H-bond–forming atoms. Equation (S12) shows an example of protein-ligand H-bond competing process, in which two ligand atoms L_1 and L_2 compete H-bonding with a H-bond donor (P-H) from protein. All H-bonds are expressed with dashed lines. Equation (S12) is decomposed into two parts as shown in the box. R_1, R_2, r_1 and r_2 stand for the H-bond distances between H and L_1 or L_2. k is q_2/q_1, where q_2 and q_1 are the charges of L_2 and L_1 . E_C stands for the electrostatic interaction between P-H and L_1. E'_C stands for the electrostatic interaction between HO-H and L_1.

$$P\text{-}H\text{---}L_1 + HO\text{-}H\text{---}L_2 \leftrightarrow P\text{-}H\text{---}L_2 + HO\text{-}H\text{---}L_1 \quad (S12)$$



The enthalpy change (△H) for equation (S12) can be calculated based on equations (S5) and (S6) from this Supplementary Information. Because large change of electrostatic interaction causes only small change of H-bond distance, we cancel the second term for equations (S5) and (S6) for simplification. Thus the enthalpy change for equation (S12) can be expressed as equation (S13). We also assume $r_1/r_2 \approx R_1/R_2$ because H-bond distance change from r_1 to r_2 is similar to the H-bond distance change from R_1 to R_2 and large change of electrostatic interaction causes only small change of H-bond distance. Thus, equation (S14) can be obtained from equation (S13) when $r_1/r_2 = R_1/R_2$.

$$\Delta H = \frac{11}{12} \times E'_c \times \left(\frac{k \times R_1}{R_2} - 1\right) - \frac{11}{12} \times E_c \times \left(\frac{k \times r_1}{r_2} - 1\right) \quad (S13)$$

$$\Delta H = \frac{11}{12} \times (E_c' - E_c)\left(\frac{k \times R_1}{R_2} - 1\right) \quad \text{(S14)}$$

E_C and E'_C are calculated based on Coulomb's law. Assume the charges for the hydrogen atom and oxygen atom in water are q_{WH} and q_{WO}, then equation (S15) is obtained from equation (S14) and Coulomb's law, where c equals to $11/(48*\pi*\varepsilon_0)$. Equation (S17) can be obtained equation (S16), which is in turn obtained from equation (S15).

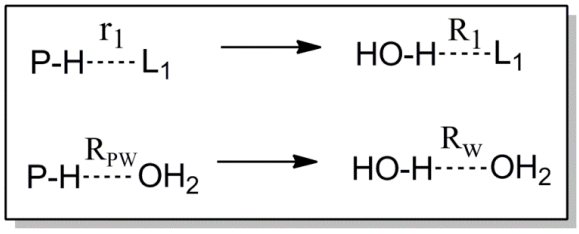
$$\Delta H = c \times \left(\frac{q_1 \times q_{WH}}{R_1} - \frac{q_1 \times q_{PH}}{r_1}\right)\left(\frac{q_2 \times R_1}{q_1 \times R_2} - 1\right) \quad \text{(S15)}$$

$$\Delta H = \frac{c \times R_1}{q_{WO} \times q_{WH}} \times \left(\frac{q_{WO} \times q_{WH}}{R_1} - \frac{q_{WO} \times q_{PH}}{r_1}\right)\left(\frac{q_{WH} \times q_2}{R_2} - \frac{q_{WH} \times q_1}{R_1}\right) \quad \text{(S16)}$$

$$\Delta H = \frac{c \times R_w}{q_{WO} \times q_{WH}} \times \left(\frac{q_{WO} \times q_{WH}}{R_w} - f_r \times \frac{q_{WO} \times q_{PH}}{R_{PW}}\right)\left(\frac{q_{WH} \times q_2}{R_2} - \frac{q_{WH} \times q_1}{R_1}\right) \quad \text{(S17)}$$

In equation (S17), f_r is close to 1 as the H-bond distance change from r_1 to R_1 is similar to the H-bond distance change from R_{PW} to R_W as shown in the following box.

$$f_r = \left(\frac{R_1}{r_1}\right) \Big/ \left(\frac{R_W}{R_{PW}}\right) = 1$$



Assume E_{WH_W} is the electrostatic interaction between the hydrogen atom of water and the oxygen atom of water; E_{PH_W} is the electrostatic interaction between P-H and the oxygen atom of water; E_{L1_W} is the electrostatic interaction between L_1 and the hydrogen atom of water; E_{L2_W} is the electrostatic interaction between L_2 and the hydrogen atom of water.

Thus: $E_{WH\_W} = q_{WO}*q_{WH}/(4*\pi*\varepsilon_{0*}R_W)$ (S18)

$$E_{PH\_W} = q_{WO} * q_{PH} / (4 * \pi * \varepsilon_0 * R_{PW}) \quad \text{(S19)}$$

$$E_{L1\_W} = q_{WH} * q_1 / (4 * \pi * \varepsilon_0 * R_W) \quad \text{(S20)}$$

$$E_{L2\_W} = q_{WH} * q_2 / (4 * \pi * \varepsilon_0 * R_{PW}) \quad \text{(S21)}$$

By substituting equations (S18), (S19), (S20) and (S21) into equation (S17), we obtain equation (S22), in which c' is a constant.

$$\triangle H = c' * (E_{WH\_W} - E_{PH\_W}) * (E_{L1\_W} - E_{L2\_W}) \quad \text{(S22)}$$

As H-bonding capability of an atom is the free energy change from water to hexadecane by breaking its interaction with water, the relationship between $\triangle G$ and H-bonding capability is similar to the relationship between $\triangle H$ and electrostatic interaction with water. Thus equation (S23) is obtained based on equation (S22).

$$\triangle G \approx c'' * (H_{WH} - H_{PH}) * (H_2 - H_1) \quad \text{(S23)}$$

where c'' is a constant, H_{WH} , H_{PH}, H_1 and H_2 are the H-bonding capabilities for the Hydrogen of water, P-H, L_1 and L_2 respectively.

Thus, the $\triangle G$ for equation (S12) can be expressed with equation (S24) or equation (S25).

$$\mathbf{\triangle G \propto (H_{WH} - H_{PH}) * (H_2 - H_1)} \quad \textbf{(S24)}$$

$$\text{or: } \mathbf{\triangle G = k * (H_{WH} - H_{PH}) * (H_2 - H_1)} \quad \textbf{(S25)}$$

Validation:

(i) Assume the environment in protein is similar to that in hexadecane(H_{PH} = 0), the $\triangle G$ for transferring L_1 and L_2 from water to protein are the same to those for transferring L_1 and L_2 from water to hexadecane, which are H-bonding capabilities of L_1 and L_2. Thus the $\triangle G$ for equation (S12) is H_2 - H_1, which is in agreement with equation (S25) and k equals 1/ H_{WH} (please see Fig. 2 for details).

If the H-bonding capability of PH is the same to that of the hydrogen atom in water, the free energy change for transferring L_1 and L_2 from water to protein are zero, and the free energy change for equation (S12) is zero, which is in agreement with equation (S25).

Thus, no matter what L_1 and L_2 are, equations S24 and S25 are always correct when the H_{PH} equals to zero or H_{WH}. Although the derivation assumes that there is special relationship between the H-bond distances, equations S24 and S25 are still correct if the H-bond distances don't have the relationship because the H-bonding capabilities of L_1 and L_2 can be any values. Thus, the approximations in the derivation don't affect the accuracy of equations S24 and S25.

(ii) if $H_2 = Hw$ and $H_1 = 0$ (this is: L_1 is non-polar atom, L_2 has the same H-bonding capability as water),

$$P\text{-}H\text{---}L_1 \rightarrow P\text{-}H\text{---}L_2 \qquad \triangle G_1 = -H_{PH} \qquad \text{(S26)}$$

$$HO\text{-}H\text{---}L_2 \rightarrow HO\text{-}H\text{---}L_1 \qquad \triangle G_2 = H_{WH} \qquad \text{(S27)}$$

the $\triangle G$ for equation (S12) is the sum of the $\triangle G$ of equations (S26) and (S27), which equals to $H_{WH} - H_{PH}$, which is in agreement with equation (S25) and k equals $1/ H_{WH}$.

Thus, no matter what H_{PH} is, equations (S24) and (S25) are always correct when $H_2 = Hw$ and $H_1 = 0$.

Text S3. The H-bonding capability of the protein atoms with which a ligand atom interacts and the effect of H-bond geometry on the H-bond interaction. For a ligand atom, the H-bonding capability of the protein atoms (H_P) is estimated based on the following format:

$$H_P = \sum(H_i * f_d) \quad (1 \geq f_d \geq 0) \tag{S26}$$

H_i is the H-bonding capability of a protein atom i. f_d is a correction factor depending on the distance (d) between the ligand atom and protein atom i. For simplification, we roughly set f_d to be 1 if the d is 1.8Å or less because the optimal distance for the H-bonds in the water is 1.8Å. We assume f_d to be 0 if d larger than 3.2Å, which is twice the van der Waals radii of the H-bond acceptor because there is equal possibility for attraction interaction and repulsion interaction and the total electrostatic interaction would be very low. Based on the relationship between electrostatic interactions (E_c) and the H-bond distance d ($E_c \propto 1/d$) and the effect of H-bond distance on non-covalent van der Waals interactions (E_{vdw}), the relationship between f_d and d is roughly expressed with equations (S27) and (S28).

$$f_d = 0.0833 + 1.65/d \quad (1.8 < d < 2.6) \tag{S27}$$

$$f_d = 0.72 * (1- (d-2.6)/0.6) \quad (2.6 < d < 3.2) \tag{S28}$$

The following table lists the fd values for various distances between H-bond donor and acceptor:

D	1.8	1.9	2.0	2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8	2.9	3.0	3.1	3.2
f_d	1.0	0.95	0.91	0.87	0.83	0.80	0.77	0.74	0.72	0.60	0.48	0.36	0.23	0.12	0

In this study, we assume f_d to be 1 if d is not much different from 1.8Å for the following reasons: i) the coordinate uncertainties in crystal structures of proteins always exist; ii) atoms can move by rotating the rotatable bonds so that H-bond donors and acceptors try to interact at the optimal distance; iii) The favorable electrostatic interaction decreases as d increases, but the unfavorable non-covalent Van der Waals (E_{vdw}) also decreases largely if the H-bond distance is around the optimal distance.

In this study, we only consider the H-bonds with H-bond angles that don't deviate largely from the ideal geometries. Also, we don't consider the effects the H-bond angles on the free energy contributions of

the H-bonds because the H-bond with a broad distribution of Angles is more favorable than the well-oriented fix H-bond with ideal geometry in free energy. This can be illustrated by the fact that the H_2O…HOH H-bonds with ideal geometries in ice are much stronger than those in water, but water is more stable than ice at room temperature.

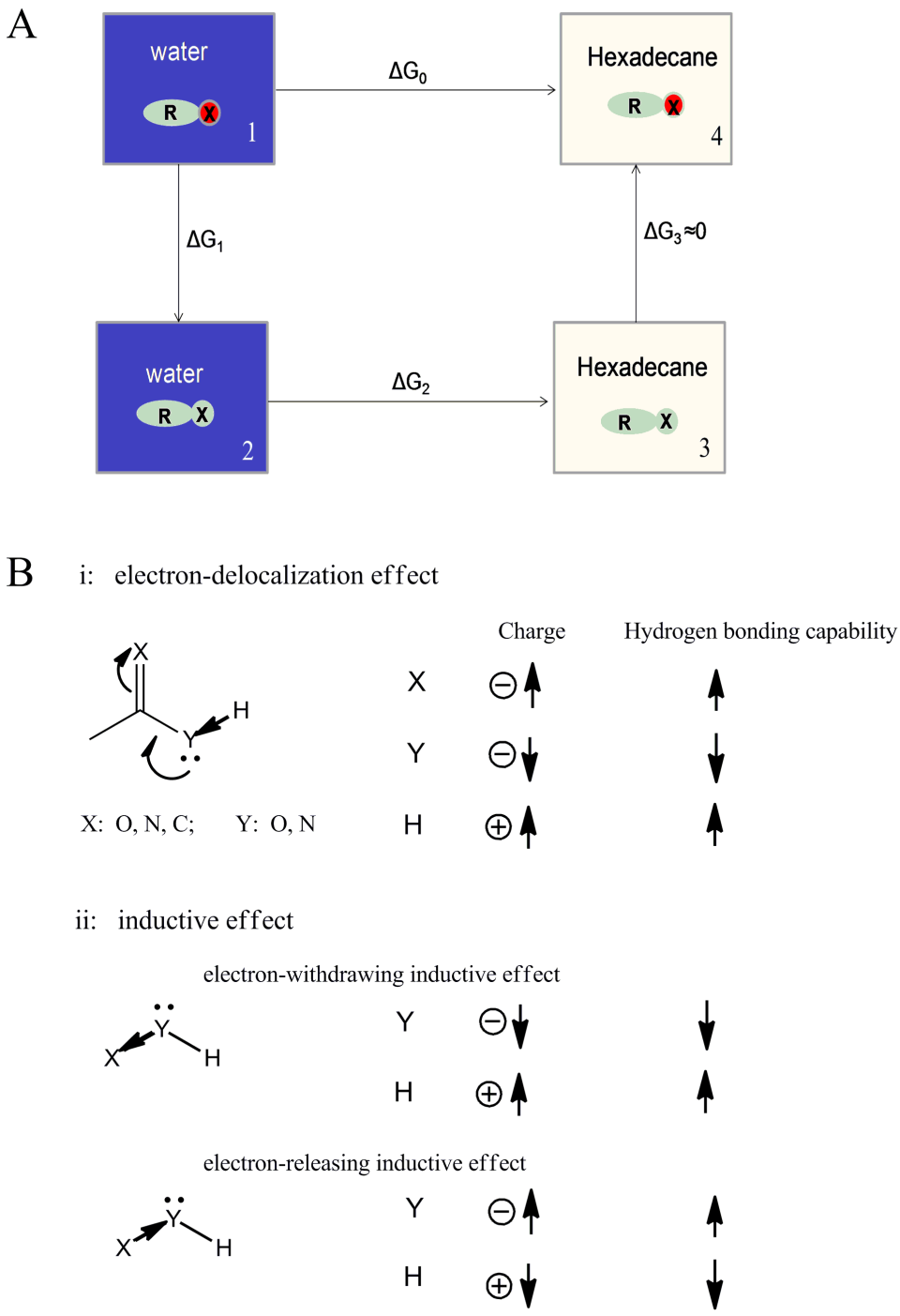


fig. S2. Calculation of H-bonding capability based on water/hexadecane partition coefficients. The H-bonding capability of an atom is estimated from the thermodynamic cycle (shown in **A**), where RX is a small molecule consisting of R (alkyl) and X. Red and grey represent chemical groups able or unable of forming H-bonds, respectively. We define the H-bonding capability of X as the free energy change for the process from State 1, in which X can form an H-bond with water, to State 2, in which X is

unable to form H-bonds with water. That is: the H-bond capability of X is ΔG_1. Based on the thermodynamic cycle, ΔG_1 can be calculated by the following equation:

$$\Delta G_1 = \Delta G_0 - \Delta G_2 - \Delta G_3$$

ΔG_0 represents the free energy required to transfer RX (red) from water to hexadecane, calculated from its water-hexadecane partition coefficient. ΔG_2 represents the free energy required to transfer the RX (grey) from water to hexadecane. Because ΔG_2 does not involve H-bond formation with water, it is calculated from its solvent (water) accessible surface area. Because no H-bonds are formed with hexadecane, ΔG_3 is close to zero and is ignored. Therefore, the H-bonding capability of X in RX is calculated from the water-hexadecane partition coefficient of RX (for ΔG_0) and the water accessible surface area of RX (for ΔG_2). Because most molecules contain more than one H-bond forming atom, the H-bonding capability of each atom is calculated using QSAR models, based on water-hexadecane and water-octanol partition coefficients (Materials and Methods in this Supplementary Information). Table 1 shows the H-bonding capabilities of individual atoms in common functional groups calculated using this method.

Electrons can exert significant influence on H-bonding capabilities as illustrated in B. These include: *i)* resonance effects where electron redistribution of X, Y, and H is caused by delocalization through interconnected π-bonds, and ii) inductive effects where redistribution of electron density of Y and H occurs through σ bonds. Increasing the electron density of an H-bond acceptor or decreasing the electron density of an H-bond donor will both increase their H-bonding capability. B (*i*) shows that delocalization of lone pair electrons on Y to the π double bond increases the negative charge and H-bonding capability of X; decreases the negative charge and H-bonding capability of Y; and increases the positive charge and H-bonding capability of H. B (*ii; top illustration*) shows that electron withdrawing groups decrease the H-bonding capabilities of H-bond acceptors (HBA) and increase the H-bonding capabilities of H-bond donors (HBD). B (*ii; bottom illustration*) shows electron releasing groups increase the H-bonding capabilities of HBA and decrease the H-bonding capabilities of HBD. For

example, alkyl is an electron-releasing group. The oxygen atom of alcohol has a higher H-bonding capability than the oxygen of water, while the hydrogen atom of alcohol has a weaker H-bonding capability than the hydrogen of water. The H-bonding capability of an HBA with lone pair electrons is strongly related to its electronegativity. For example, because nitrogen is less electronegative than oxygen, ammonia is a much stronger HBA than water. One the other hand, N-H compounds are much weaker HBD than the corresponding O-H compounds (Table 1).

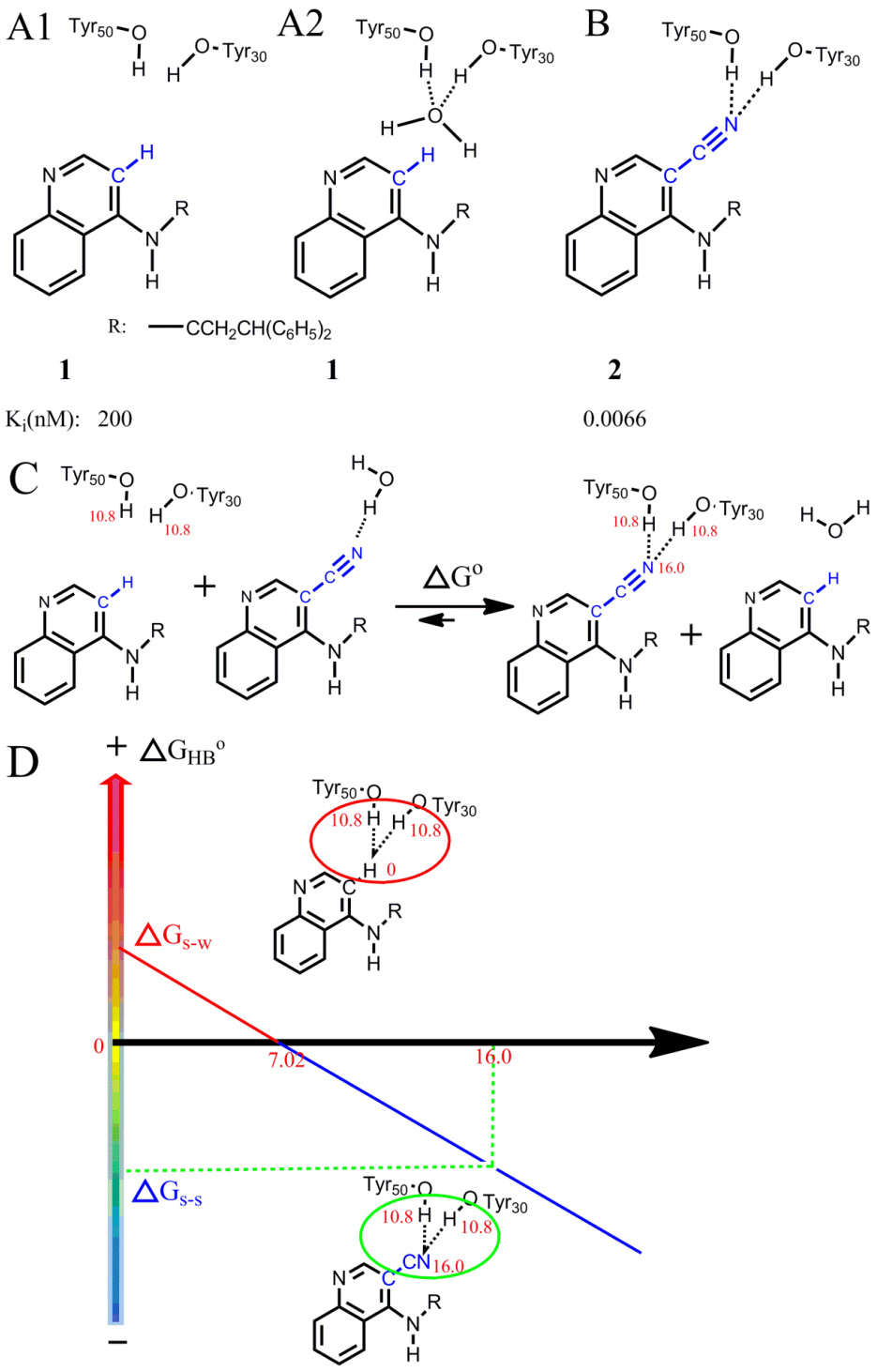


fig. S3: Contributions of the H-bonds between CN and the Tyr-OH from scytalone dehydratase to protein-ligand interactions. (**A,B**) The interaction modes of two scytalone dehydratase inhibitors **1** and **2**. It was reported that there is a water molecule between CH of **1** and Tyr-OH of protein (as in Figure S3A2) (26). But the analysis based on similar crystal structure (e.g. 3STD, 5STD) indicates that there is no enough space for holding a water molecule. Moreover, the much higher binding affinity of 2 doesn't

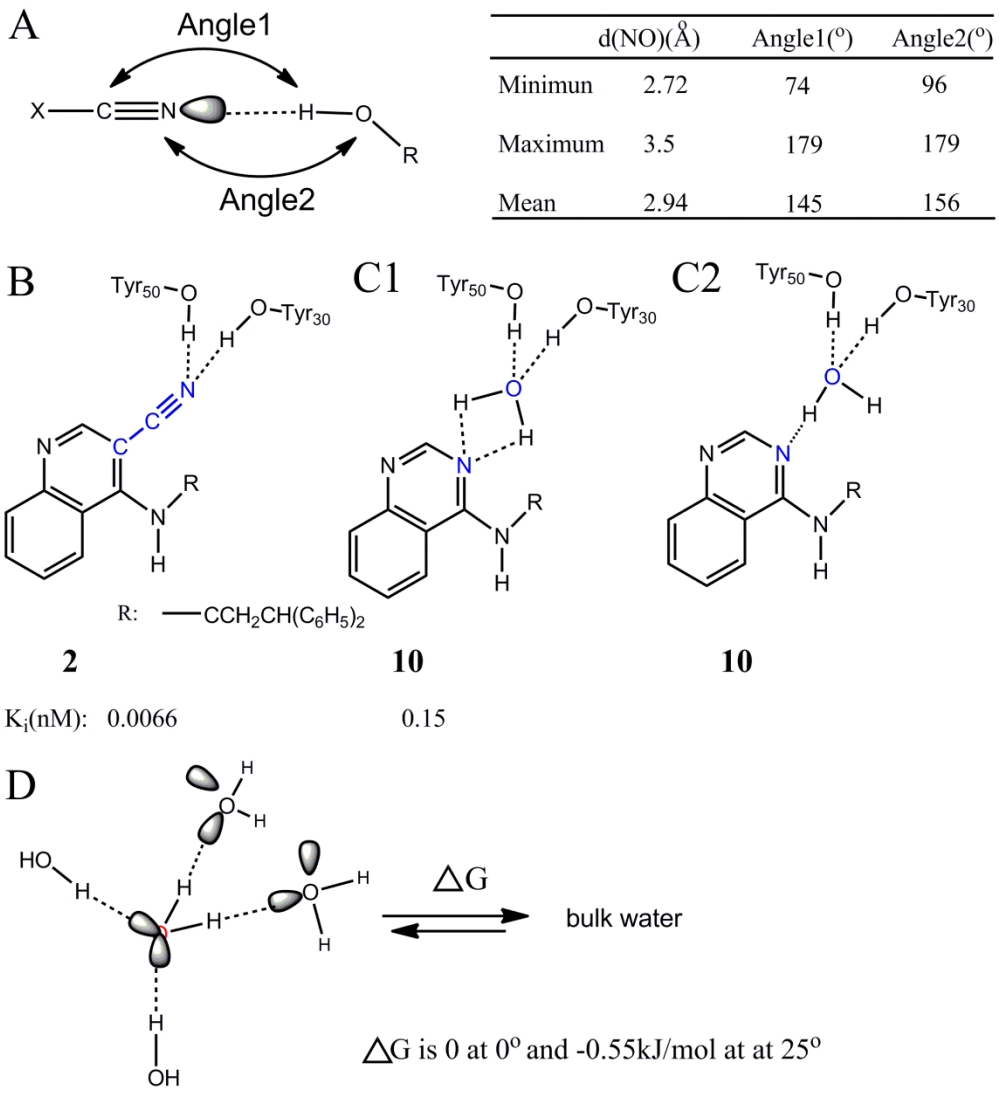
result from replacing the water molecule into bulk water, which is proved in Figure S4. Thus, the binding mode of **1** is A1, not A2. (**C**) The H-bond compete paring process between **2** and **1** (with mode A1). The equilibrium constant for the process (K) is: 200/0.0066 = 30,300. The standard free energy change for the process is: $\Delta G^\circ = -RT\ln K = -25.57 kJ/mol$. (**D**) The relationship between the H-bonding capability for the ligand atom interacting with Tyr-OH and the free energy contribution of its H-bond interactions to protein-ligand interaction.

$$\Delta G^\circ = \Delta G_{s-s} - \Delta G_{s-w} = -25.57 kJ/mol$$

Based on $\Delta G_{HB} = k \times (H_W - H_P) \times (H_L - H_W)$: we get:

$$\frac{\Delta G_{s-s}}{\Delta G_{s-w}} = \frac{16.0-7.02}{0.0-7.02}$$

Thus: $\frac{\Delta G_{s-s}}{\Delta G^\circ} = \frac{16.0-7.02}{16.0-0}$; and $\frac{\Delta G_{s-w}}{\Delta G^\circ} = \frac{0.0-7.02}{16.0-0}$. Then, we get: ΔG_{s-s} = -14.35kJ/mol, indicating that the s-s pairing H-bonds increase ligand binding affinity about 329-fold; ΔG_{s-w} = 11.22kJ/mol, indicating the s-w pairing interaction decrease ligand binding affinity about 93-fold.



	d(NO)(Å)	Angle1(°)	Angle2(°)
Minimun	2.72	74	96
Maximum	3.5	179	179
Mean	2.94	145	156

fig. S4: Proof for the strong H-bond interactions between the CN group of inhibitor 2 and tyrosine hydroxyls from scytalone dehydratase. It was reported that the interactions between the CN group of inhibitor 2 and tyrosine hydroxyls are weak interactions between nitrile π-electrons and the tyrosine hydroxyls because the H-bond angles from the nitrile to the tyrosine hydroxyls cannot be 180° angle (26). Here, we prove that there is strong H-bond interactions the CN group of inhibitor 2 and tyrosine hydroxyls from scytalone dehydratase.

Proof 1: Theoretical studies and gas-phase experimental structures have both indicated that the sp nitrogen lone pair is the hydrogen-bond donor (49).

Proof 2: The geometric parameters describing H-bonding interactions on the nitrogen atom of nitriles obtained from Cambridge Structural Database showing in Figure S4A (adopted from (49)) indicates that the sp lone pairing can adopt broad distribution of conformations. For the interaction between the CN group of inhibitor **2** and the two Tyr-OH groups of scytalone dehydratase, the distances between N and O [d(NO)] are 2.8~2.9Å; Angle1 ranges from 109~138° and Angle2 ranges from 130~170° angle. The d(NO) values are close to minimum value (Figure S4A) , while the angles are not close to the maximum values. However, the H-bond with a broad distribution of angles is more favorable than the well-oriented fix H-bond in free energy. This can be illustrated by the fact that the H_2O..HOH H-bonds with ideal geometries in ice are much stronger than those in water, but water is more stable than ice at room temperature. Thus, the H-bond interactions between the CN group of inhibitor **2** and the two Tyr-OH groups of scytalone dehydratase are strong interactions.

Proof 3: If the Tyr-OH forms weak interactions with pi-electrons, it is impossible to explain why the inhibitor **2** is over 20-fold more active than inhibitor **10**, which can form four (Figure S4C1) or three (Figure S4C2) strong H-bonds. To compare the H-bonds of **2** with **10**, the transferring process from ice to water (Figure S4D) is used to estimate the free energy change for releasing constrained water to bulk water. The free energy change for the process at 25°C is -0.55kJ per mol H_2O (=-0.001*1mol*25K*22J mol^{-1} K^{-1}), which means the free energy change to break two mols of well oriented H_2O…HOH H-bonds and then release one mol of constrained water to bulk water is -0.55kJ. If the H-bonds between CN and Tyr-OH (Figure S4B) are weak, inhibitor **10** has three or four strong H-bonds (stronger than H_2O…HOH H-bond) more than inhibitor **2** and inhibitor **10** must be much more active than inhibitor **2**, which is disagree with the experimental data. Thus, the H-bonds between CN and Tyr-OH (Figure S4B) are strong.

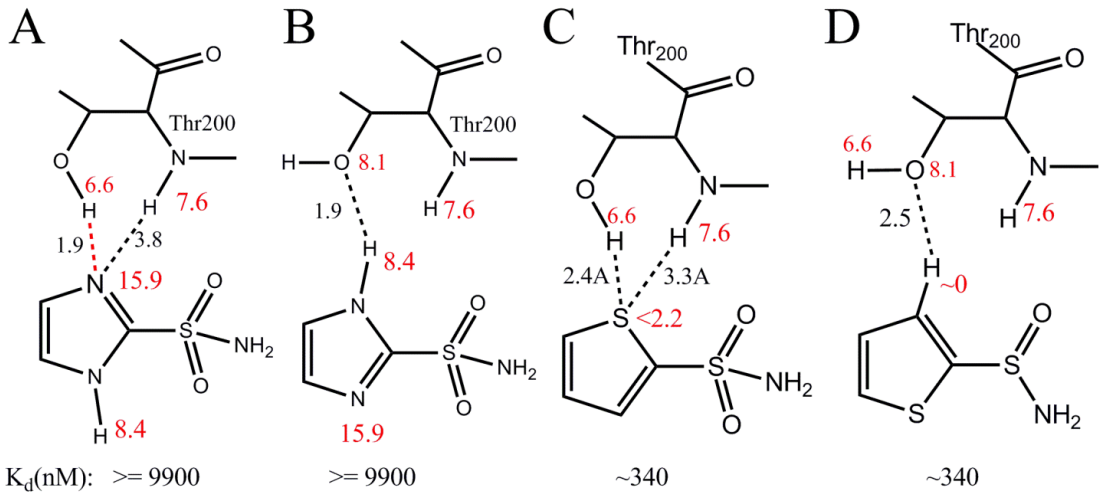


fig. S5. Binding affinities of 1*H*-imidazole-2-sulfonamide and thiophene-2-sulfonamide. The binding affinities of 1H-imidazole-2-sulfonamide and thiophene-2-sulfonamide to carbonic anhydrase (31) are 9,900nM (K_d) and 340 nM (K_d) respectively. There are two possible binding modes for 1H-imidazole-2-sulfonamide (**A&B**). The H-bond which is colored red in (**A**) is much stronger than the H-bonds in water because the H-bond acceptor (N) has much stronger H-bonding capability than that of water and the H-bond donor has H-bonding capability close to the H atom of water. The crystal structure for the thiophene-2-sulfonamide complex (pdb code: 3S78) indicates that there are two possible binding modes for thiophene-2-sulfonamide(**C&D**), indicating that the two binding modes have similar stabilities and the strengths of the H-bonds in (**C**) are close to 0.

We are interested in how the H-bond colored red in A contributes to ligand binding affinity. No matter how 1H-imidazole-2-sulfonamide interacts with carbonic anhydrase in mode A or in mode B, the K_d for mode A does not reduce below 9,900 nM.

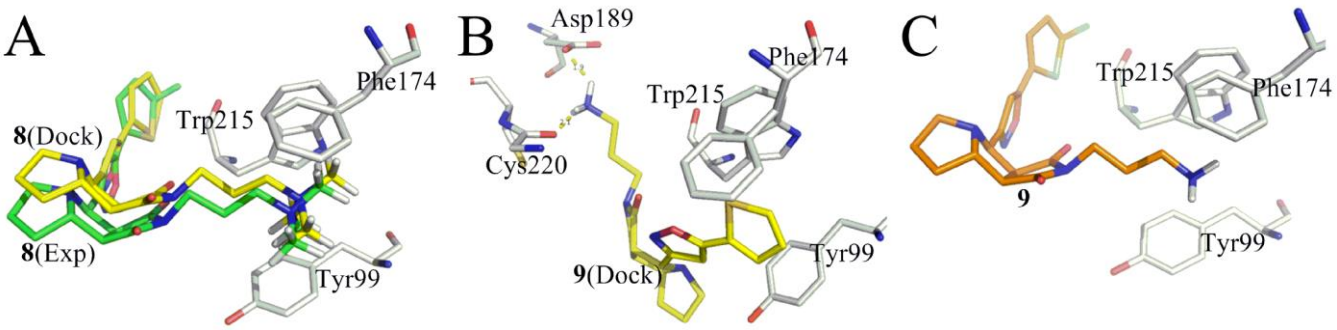


fig. S6. Quaternary ammonium cation [–N(Me)$_3^+$]-π interactions are more favorable than ammonium ion (−NH$_3^+$)-π interactions. (**A**) Crystal structure of factor Xa and **8** (colored green in **A**) complex (pdb code: 2JKH) and the docking mode (colored yellow). Both show that the –N(Me)$_3^+$ group interacts with the aromatic rings. (**B**) The docking mode of **9** shows that the -NH$_3^+$ group does not interact with the aromatic rings, but with Asp189 and Cys200. (**C**) The imaginary interaction mode of **9** which is the same as the interaction mode of **8**.

If the interaction mode of **9** was the mode shown in (**C**), the binding free energy difference between **8** and **9** originated the difference between –N(Me)$_3^+$-π interactions and -NH$_3^+$-π interactions. But the actual interaction mode of 9 is not the mode as shown in (**C**), indicating the activity of **9** would be lower if **9** adopts the mode shown in (**C**). Thus, –N(Me)$_3^+$-π interactions is more than 17.3kJ/mol, which is the binding free energy difference between **8** and **9**, favorable than the -NH$_3^+$-π interactions.

The large difference cannot be explained by the van der Waals interactions between –N(Me)$_3^+$ and the aromatic rings because similar van der Waals interactions, e.g. the interactions with solvent, exist before binding. The SASA of **8** is about 57Å^2 larger than that of **9**, which corresponds to 9.6kJ/mol of desolvation energy calculated from equation (6) in the experimental section, much less than the binding free energy difference. Thus, the favorable quaternary ammonium cation-π interactions cannot be explained merely with desolvation energy or van der Waals.

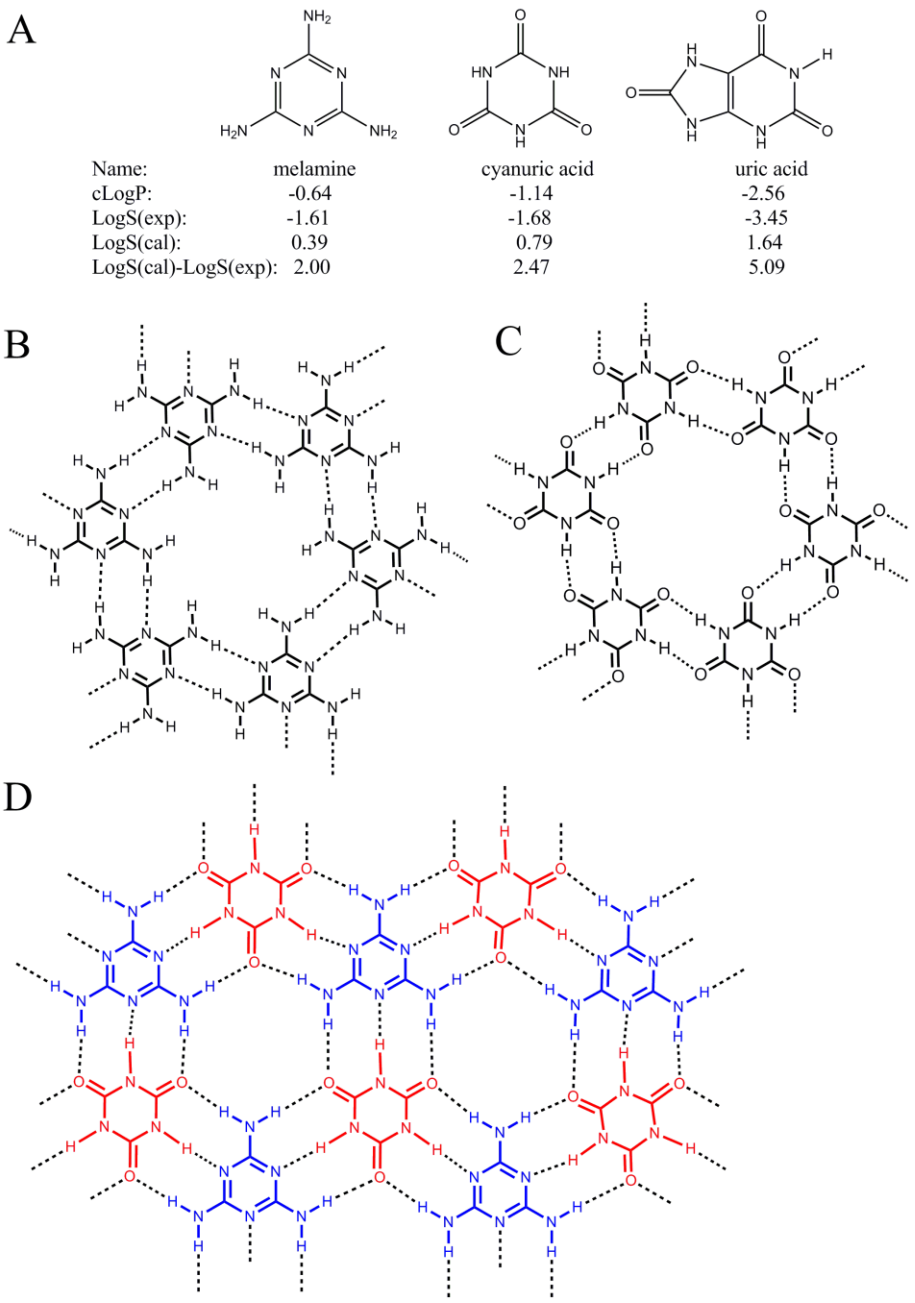


fig. S7. A pathogenic role for the s-s/w-w H-bonding principle in melanin toxicity. Melamine toxicity has become widely publicized after recent occurrences of renal injury in infants and children exposed to melamine-tainted milk in China. This renal damage is believed to result from kidney stones formed from melamine and uric acid or from melamine and its co-crystallizing chemical derivative, cyanuric acid (37,50). The formation of such crystals (kidney stones) can be explained by the s-s H-

bond pairing principle when reviewing the structures of the 3 compounds and their intermolecular H-bond interactions. **A**) Structures of the 3 compounds and their solubility shown as LogS(exp), the log value of experimental water solubility in mol L^{-1}. logS(cal) is the value calculated from an equation developed by Meylan (51): log *S* (mol L^{-1}) = 0.920 – 0.834 log*P*– 0.0084 MWT (n = 1450; r^2 = 0.875; sd = 0.804), where MWT represents molecular weight. These compounds have low water solubility and a negative clogP. Usually, compounds with a negative logP have a good water solubility. However, the actual water solubility is much lower than the calculated value, implying that the intermolecular interactions are much stronger than expected. The strong intermolecular interactions of those compounds result from intermolecular s-s H-bond pairings, supported by the experimental finding that melamine and cyanuric acid dissolve in acetonitrile and pyridine (52). Acetonitrile and pyridine contain HBAs that have much stronger H-bonding capabilities (18.1 and 18.2 respectively) than water (7.02), and forms preferable s-s H-bond pairings that disrupts the melamine-cyanuric acid complex. This principle is demonstrated experimentally in Figure 7. (**B** & **C**) Intermolecular H-bonds for melamine and cyanuric acid, respectively; (**D**) melamine and cyanuric acid form highly organized intermolecular H-bond pairings. The hydrogen and nitrogen atoms on the aromatic rings have stronger H-bonding capabilities than those for the hydrogen and oxygen of water and therefore form strong-strong pairings, resulting in poor complex solubility in water.

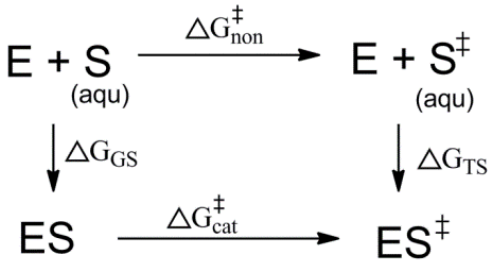


fig. S8: The thermodynamic cycle that demonstrate the contribution of H-bonds to enzymatic catalytic power equates to the contribution to protein-ligand binding: S and $S^{\neq}$ stand for the ground state and transition state of a reaction in aqueous solution; E stands for the enzyme catalyzing the reaction. The free energy barriers for the solution reaction and enzymatic reaction are $\Delta G^{\neq}_{non}$ and $\Delta G^{\neq}_{cat}$ respectively. The free energy barrier difference between the enzymatic reaction and its reference reaction ($\Delta G^{\neq}_{cat}$ - $\Delta G^{\neq}_{non}$) equals to the substrate-enzyme binding free energy in transition state (ΔG_{TS}) minus that in ground state (ΔG_{GS}). For a newly formed H-bond in transition state, the contribution of the H-bond to the reduction of free energy barrier of the enzymatic reaction equates to the contribution of the H-bond to substrate-enzyme binding in transition state.