DRAFT

A TEMPLATE-BASED APPROACH TO THE MODIFICATION OF BINDING PROPERTIES OF GLOBULAR PROTEINS II: *Rationale and proposed approach.*

Imadol V. Jeff-Eke[1]

Abstract

We present a rationale and proposed approach to the modification and development of bind sites using their respective cognate ligands as template. This is in support of a plausible "instructive" role for the ligand and therefore its involvement in determination of the structure and properties of bind sites. We emphasize the relationship between substrate and active site as an example of the relationship between ligand and bind sites, respectively. This is based on the assumption that there are shared features between all ligand:bind site complexes. Therefore, principles that apply to a specific complex can be applied, in general, to other protein-based complexes. We define ligand-associated probability bias as the difference between the probability of finding activity-determining conformations (ADCs) in the presence- and absence of ligands. For cognate ligands, the given bias is in favor of these ADCs. Thus, bind sites are more likely to assume ADCs when their cognate ligands are present. We relate such probability bias to structural reorganization, disorganization, and preorganization events. We then propose a means of deriving an [apparent] preorganized bind site structure by way of reorganization events that occur with cognate ligand. Finally, we propose a means of deriving an [actual] preorganized bind site structure by way of reorganization events that occur with cognate ligand, albeit during the folding process. The assumption is that the role of the ligand in derivation of such [actual] preorganized bind site structures is an instructive role, and is in support of the Haurowitz-Pauling hypothesis.

Keywords: Bind sites; active sites; cognate ligands; substrates; conformational selection; induced fit; conformational transitions; reorganization; disorganization; preorganization.

Ligand-bind site interactions and complementarity

Consider that all molecular activities known to occur in biological systems and involving two separate entities of which, at least, one is a polypeptide [substrate-enzyme, antigen-antibody (epitope-

[1] Contact information: ivjeffeke@gmail.com

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2161v2 | CC BY 4.0 Open Access | rec: 6 Jul 2016, publ: 6 Jul 2016

paratope), allosteric modulator-allosteric site, receptor ligand-receptor, and other peptide-peptide interactions] must be preceded by physical interactions between involved entities, in a binding process. A requirement for such binding is that the involved surfaces of these entities complement one another: either by docking of the ligand within a bind site groove in the form of **structural complementarity;** or optimal pairing of bind site dipoles with charged groups on ligand in the form of **electrostatic complementarity**[1]; or a combination of both types of complementarity. Irrespective of which, we suppose that an equilibrium involving ligand:peptide binding occurs whenever a free polypeptide, **[P]**, and a cognate binding ligand, [L], are introduced.

$$[\mathbf{P}] \; + \; [\mathrm{L}] \leftrightarrow [\mathbf{P}\mathrm{L}]$$

Although seemingly obvious, we shall however mention that formation of the ligand:peptide complex, **[PL]**, would occur in an absence of external manipulation. This is especially important for experimental protocols involving mixing of ligand and peptide. Reasons for such occurrence are not considered in this paper. A very important point of mention is that although variations in binding phenomena may exist for different types of ligand:bind site complexes, we suppose that general themes do exist for the mechanics of these processes. Thus, at least in terms of these shared features, we can apply principles conceived for a given category of ligand:bind site complex formation to other categories. For example, those physical and chemical principles that govern formation of substrate:enzyme complexes should also apply to formation of antigen:antibody complexes and vice versa. We shall focus on substrate:enzyme complexes and catalytic activities that may derive from them. We present an argument in support of a potential "instructive" role for the ligand (substrate) and therefore its involvement in determination of bind site structure and properties.

Probability of finding a given bind site conformation on a physiologically folded peptide[2] under equilibrium conditions and in an absence of ligand.

Two models, induced and selection models, are of importance here. Although they differ from the Haurowitz-Pauling hypothesis in that they concern physiologically folded peptides –peptides in their native states; they can be applied here since they concern ligand influences in determination of those conformations that predominate a complexed bind site. Thus, they are of relevance to this interest. We

[2] Although polypeptide and protein, as used in this work, are applied here as interchangeable terms; all amino acids that make up the structure are considered to be contained within the same primary structure. If multiple chains, then these chains should be considered linked by disulfide bridges, so as to allow a continuity of covalent chemical bonding for the structure(s).

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2161v2 | CC BY 4.0 Open Access | rec: 6 Jul 2016, publ: 6 Jul 2016

shall begin by reconciling both models into a single probability-based model so as to avoid having to produce separate arguments for each.

At a time, the prevailing explanation for enzyme:substrate complementarity was that posited by Emil Fischer, the key-lock model which holds that a binding ligand is structurally complementary to a predominant conformational state of the local region on the peptide onto which it binds –active site. Thus the lock and key model implies a [near] static bind site conformation and a predominance of structural complementarity[2]. However, experimental data were in support of a more dynamic bind site, as opposed to the static, pre-existing active site proposed by the key-lock model.

Of the models proposed to explain the dynamic nature of bind sites, two dominated discourse: induced-fit and conformation selection models. Since there are multiple implications for either model, multiple formulations can be emphasized for each. Here we emphasize two such formulations. **First formulations** for both induced-fit and conformational selection models may be considered to contemplate relationships between set(s) of conformations that occur in the presence and absence of ligand[3-6]. **Second formulations** for both models may be considered consequences of their respective initial formulations, and also intended to contemplate relationships between set(s) of conformational transitions that occur in the presence and absence of ligand

First formulation for the **induced-fit model** alludes to the notion that some bind site *conformations –of special interest are* **activity-determining conformations (ADCs)**– *t*hat occur with ligand-peptide interactions would otherwise not have occurred in an absence of the ligand[7,8]. In other words, the probability, P^o, of finding bind sites that assume these ADCs, in an absence of the ligand is: $P^o = 0$. A second formulation holds that the ligand first binds to the bind site, followed by assumption of ADCs by bind site. Thus the given conformational transition, involving transitions into these ADCs will not occur in an absence of ligand.

First formulation for the **conformation selection model** alludes to the notion that bind site ADCs that occur with ligand:peptide interactions are among the repertoire of conformations that pre-exist ligand peptide interactions and therefore would have occurred even in an absence of the ligand[8]; with the ligand merely affecting the stability of these ADCs. This effect of ligand is believed to affect the conformation equilibrium of the peptide ensemble. In other words, the probability, P^o, of finding bind sites that assume these ADCs in an absence of the ligand molecule is: $P^o > 0$. A second formulation of this model holds that the bind site first assumes ADCs, followed by ligand binding. Thus the given conformational transitions involving transitions into ADCs will occur in an absence of ligand.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2161v2 | CC BY 4.0 Open Access | rec: 6 Jul 2016, publ: 6 Jul 2016

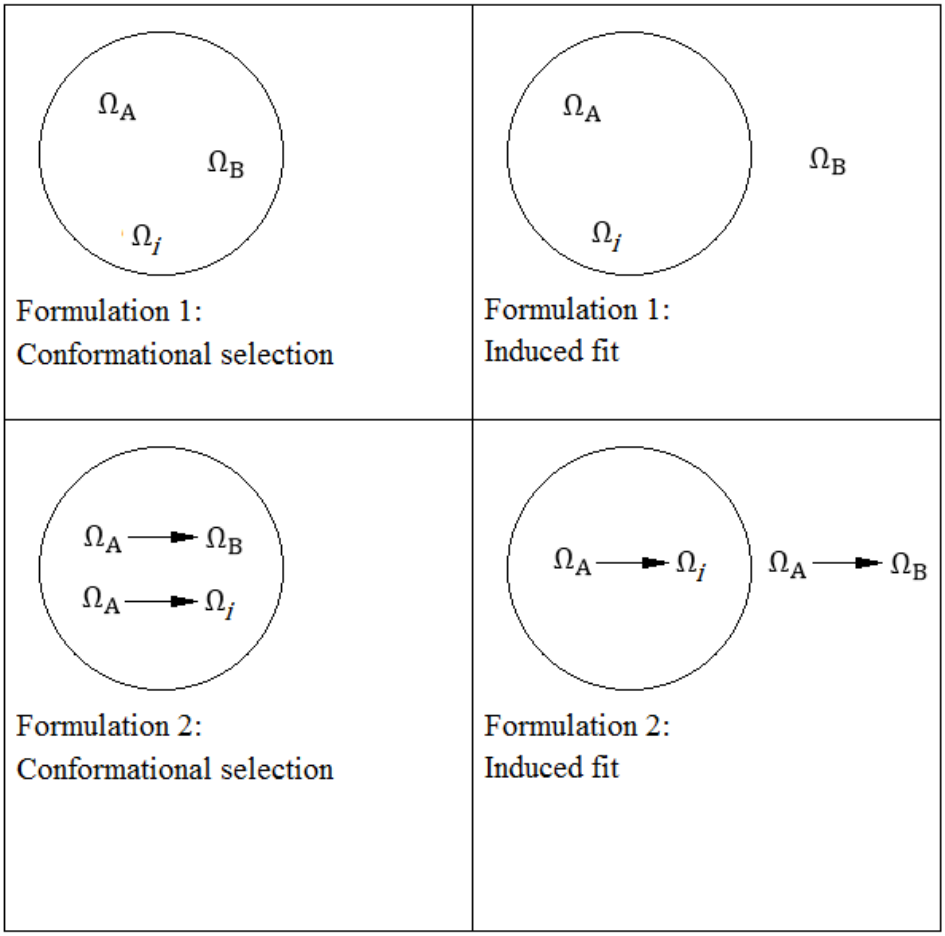

Figure 1: Illustration of each formulation for induced fit and conformation selection models. Whereas for formulation 1 all conformations are members of the set of conformations for conformation selection model, for the induced fit model, conformation B is excluded from the given set. Whereas for formulation 2 all conformational transitions are members of the set of conformational transitions for conformation selection model, for the induced fit model, conformational transition from conformation A to conformation B is excluded from the set of conformational transitions.

Determinants of "inclusiveness" within the equilibrium sets of conformations and conformational transitions

1. Let us suppose that there exists a peptide ensemble with all members of the ensemble being the given polypeptide of interest. Let us also suppose that the state of the surrounding medium wherein this occurs is such that effects of ligand on both local and global peptide conformations are negligible. Such an insignificant effect may derive from low ligand concentration or a complete absence of ligand. We shall call this the **reference condition**, and

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2161v2 | CC BY 4.0 Open Access | rec: 6 Jul 2016, publ: 6 Jul 2016

designate all parameters pertaining to this condition with a superscript (example: $P^o, \Delta E^o$, etc.). Irrespective of condition under consideration, we suppose a constant temperature for the milieu. Of the polypeptide, we are most interested in the bind/active site of which we suppose there exist only one of such bind site per peptide molecule. Thus, we can refer to the ensemble as a bind site ensemble.

2. We suppose there exists a set of conformations, members of which bind sites of ensemble map into (**Figure 2A**). A reasonable assumption is that two or more bind site(s) can map to the same conformation inasmuch as they may map to different conformations. In addition to the set of conformations, we suppose there also exists a set of conformational transitions, members of which bind sites of ensemble map into (**Figure 2B**). As was stated for bind site mapping to conformations, we assume that two or more bind sites can map to the same conformational transition. Of interest are equilibrium sets of conformations and conformational transitions.

3. Let us now suppose that we allow the bind site ensemble approach and reach conformational and conformational transition equilibria. We are interested in the probability of finding a bind site, within the given ensemble, that assumes (or maps into) a given conformation ***B*** at equilibrium.

4. Let us now apply the formulations for induced fit and conformational selection models in terms of conformation B at equilibrium. If, at equilibrium, $P_B^o = 0$ is the probability of finding a bind site that assumes a given conformation *B*, then the given conformation *B* can be said to ***not*** be included within the set of equilibrium conformations for the bind site ensemble. If at equilibrium, $P_B^o > 0$ is the probability of finding a bind site that assumes the given conformation ***B***, then the given conformation ***B*** can be said to be included within the set of equilibrium conformations for the bind site ensemble.

5. We infer from second formulations that, at equilibrium, the probability, P_B^o, of finding a bind site that assumes a given conformation B within the set of conformations is the same as the probability, P_{AB}^o, of finding a bind site that assumes a given conformational transition involving transition from an initial conformation ***A***, within the set of conformations, and transition into conformation ***B*** as opposed to a conformation ***Z*** within the set of conformations. Thus, for all practical purposes, $P_B^o = P_{AB}^o$ at equilibrium.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2161v2 | CC BY 4.0 Open Access | rec: 6 Jul 2016, publ: 6 Jul 2016

6. The question that remains to be answered is: what factors determine whether or not a given conformation is included within the equilibrium set of conformations?

7. Consider that transition from a conformation, let us call it conformation A, to a conformation, let us call this conformation *B* would depend on the number of other conformations that can also be assumed from an initial conformation *A*. Let the total number of conformations that can be assumed by direct transition from conformation A be $\boldsymbol{n}$. In addition, we suppose that at equilibrium, all transition events are equally likely and thus have equal a priori probabilities.

The probability, P_{AB}^{o}, of finding a bind site that assumes a given conformational transition involving transition from an initial conformation ***A***, within the set of conformations, and transition into conformation ***B*** as opposed to a conformation ***Z*** within the set of conformations:

$$P_{AB}^{o} = {}^{1}/_{n}$$

Also, note that the sum of all probability values must equal unity:

$$\sum_{i}^{n} P_{Ai}^{o} = \sum_{i}^{n} \left({}^{1}/_{n}\right)_{i} = 1$$

Where, $\sum_{i}^{n} P_{Ai}^{o}$ is the sum of the terms of the sequence, and the terms of the given sequence are arranged in descending order. Thus, the first and last terms of the given sequence are the maximum and minimum probability values, respectively.

$$\sum_{i}^{n} P_{Ai}^{o} = \cdots + P_{Ai}^{o} + \cdots$$

The probability, P_{AB}^{o}, can also be expressed as:

$$P_{AB}^{o} = \frac{P_{AB}^{o}}{\sum_{i}^{n} P_{Ai}^{o}} = \frac{{}^{1}/_{n}}{\sum_{i}^{n} \left({}^{1}/_{n}\right)_{i}}$$

Where, $\frac{P_{AB}^{o}}{\sum_{i}^{n} P_{Ai}^{o}}$ is a general expression and thus applies irrespective of condition; and $\frac{{}^{1}/_{n}}{\sum_{i}^{n} \left({}^{1}/_{n}\right)_{i}}$ is a specific expression that only applies to the equal a priori probability condition.

8. In addition, as n increases indefinitely, the probability, P_{AB}^{o}, approaches zero.

$$P_{AB}^{o} = \lim_{n \to \infty} \frac{{}^{1}/_{n}}{\sum_{i}^{n} \left({}^{1}/_{n}\right)_{i}} \cong 0$$

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2161v2 | CC BY 4.0 Open Access | rec: 6 Jul 2016, publ: 6 Jul 2016

From the above equality ($P_B^o = P_{AB}^o$), we can therefore conclude that, for equal *a priori* probability condition at equilibrium, the probability, P_{AB}^o, of finding a bind site that assumes a given conformation B is approximately equal to zero ($P_B^o \cong 0$) when a substantially large number of conformations occur within the set.

9. Consider that in order to transition from a given conformation A to conformation B, those interactions that characterize conformation A but not conformation B must either no longer exist or their influence on conformational transitions be negligible. Also, consider that those interactions that characterize conformation B must be formed. In other words, there should be a delta change in enthalpy of interactions, $\Delta\Delta H_{AB}^o$, for the given transition.

$$\Delta\Delta H_{AB}^o = \Delta E_{AB}^o + T\Delta\Delta S_{AB}^o$$

where,

$$\Delta E_{AB}^o = E_B^o - E_A^o$$

$$\Delta\Delta H_{AB}^o = \Delta H_B^o - \Delta H_A^o$$

$$\Delta\Delta S_{AB}^o = \Delta S_B^o - \Delta S_A^o$$

E_A^o is the free energy of conformation A or conformational energy (CE) for conformation A; E_B^o is the free energy of conformation B or conformational energy (CE) for conformation B; ΔE_{AB}^o is the free energy of conformational transition or conformational transition energy (CTE) from conformation A to conformation B; ΔH_A^o is the change in enthalpy of interaction when transitioning from a standard conformation to conformation A; ΔH_B^o is the change in enthalpy of interaction when transitioning from the standard conformation to conformation B. Note that the entropy term is a function of the number of alternate spatial arrangements for the given conformation. For our purposes, we suppose that these values are significantly diminished such that:

$$\Delta\Delta H_{AB}^o \cong \Delta E_{AB}^o$$

10. Using Boltzmann proportionality between probabilities and energy values, we offer that the probability of finding a bind site that assumes a given conformational transition (CT), within the equilibrium set of CTs for the bind site ensemble, with ΔE_{AB}^o is proportional to $exp - \left(\frac{\Delta E_{AB}^o}{kT}\right)$. Here k is Boltzmann constant and T is the absolute temperature in Kelvins.

$$P_{AB}^o \propto exp - \left(\frac{\Delta E_{AB}^o}{kT}\right)$$

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2161v2 | CC BY 4.0 Open Access | rec: 6 Jul 2016, publ: 6 Jul 2016

11. Ideally, we would want a one-to-one correspondence between a given conformational transition and a single CTE value for the given transition. That is, all CTs occupy energy states (**Figure 2**), and there are no two or more CTs that occupy the same energy state. Thus, allowing the CTE to be unique to the given CT. However, there is no evidence against the possibility that two or more different CTs may have the same CTE.

12. To attempt at approximating a one-to-one correspondence, we can define each conformational transition in terms of a range of CTEs as opposed to a single CTE values. The assumption here is that the likelihood of finding any two CTs that assume the same range of CTEs, for defined limits, would be less than if these were single CTE values. Such energy variations of a single CT may be correlated with molecular fluctuations affecting conformations involved in transition. Such thermodynamic fluctuations are noted to occur for peptide structures, thus resulting in multiple spatial arrangements of minimal variations from a most-consistent structure. Also, each of these minimally variant structural arrangements may have different energy values, thus complicating the concept of a single conformational energy value for a given peptide conformation. Thus, a variety of CTE values may result from differences between these conformational energies. To account for all such variations, we take the integral over the CTE minimum and maximum values. The probability, P^o_{AB} of finding a given conformation with CTE, is proportional to $\int_{a_B}^{b_B} exp - \left(\frac{\Delta E^o_{AB}}{kT}\right) \mathrm{d}\Delta E^o_{AB}$.

$$P^o_{AB} \propto \int_{a_B}^{b_B} exp - \left(\frac{\Delta E^o_{AB}}{kT}\right) \mathrm{d}\Delta E^o_{AB}$$

Where,

$$a_B = \overline{\Delta E}^o_{AB} - md\Delta E^o_{AB}$$

$$b_B = \overline{\Delta E}^o_{AB} + md\Delta E^o_{AB}$$

Where, $\overline{\Delta E}^o_{AB}$ is the average CTE for transition from conformation *A* to conformation *B*; and $\boldsymbol{m}$ is a non-negative integer.

13. We can express the equal a priori probabilities in terms of Boltzmann probability.

$$P^o_{AB} = {}^1\!/_n = \frac{1}{n}\frac{\left(\int_{a_B}^{b_B} exp - \left(\frac{\Delta E^o_{AB}}{kT}\right) \mathrm{d}\Delta E^o_{AB}\right)}{\left(\int_{a_B}^{b_B} exp - \left(\frac{\Delta E^o_{AB}}{kT}\right) \mathrm{d}\Delta E^o_{AB}\right)}$$

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2161v2 | CC BY 4.0 Open Access | rec: 6 Jul 2016, publ: 6 Jul 2016

$$= \frac{\int_{a_B}^{b_B} exp - \left(\frac{\Delta E_{AB}^o}{kT}\right) \mathrm{d}\Delta E_{AB}^o}{n \int_{a_B}^{b_B} exp - \left(\frac{\Delta E_{AB}^o}{kT}\right) \mathrm{d}\Delta E_{AB}^o}$$

Since all conformations have equal probabilities,

$$\int_{a_B}^{b_B} exp - \left(\frac{\Delta E_{AB}^o}{kT}\right) \mathrm{d}\Delta E_{AB}^o = \int_{a_i}^{b_i} exp - \left(\frac{\Delta E_{Ai}^o}{kT}\right) \mathrm{d}\Delta E_{Ai}^o$$

$$P_{AB}^o = \frac{\int_{a_B}^{b_B} exp - \left(\frac{\Delta E_{AB}^o}{kT}\right) \mathrm{d}\Delta E_{AB}^o}{n \int_{a_i}^{b_i} exp - \left(\frac{\Delta E_{Ai}^o}{kT}\right) \mathrm{d}\Delta E_{Ai}^o}$$

$$= \frac{\int_{a_B}^{b_B} exp - \left(\frac{\Delta E_{AB}^o}{kT}\right) \mathrm{d}\Delta E_{AB}^o}{\sum_i^n \int_{a_i}^{b_i} exp - \left(\frac{\Delta E_{Ai}^o}{kT}\right) \mathrm{d}\Delta E_{Ai}^o}$$

Where,

$a_i = \overline{\Delta E}_{Ai}^o - md\Delta E_{Ai}^o$

$b_B = \overline{\Delta E}_{Ai}^o + md\Delta E_{Ai}^o$

Where, $\overline{\Delta E}_{Ai}^o$ is the average CTE for transition from conformation *A* to a conformation for which the probability is at the i^{th} position of the pre-arranged sequence of conformational transition probabilities; and m is a non-negative integer**.** The following applies for the equal a priori probability condition and thus can be expressed as:

$$P_{AB}^o = \frac{1/n}{\sum_i^n (1/n)_i}$$

$$= \frac{\int_{a_B}^{b_B} exp - \left(\frac{\Delta E_{AB}^o}{kT}\right) \mathrm{d}\Delta E_{AB}^o}{\sum_i^n \int_{a_i}^{b_i} exp - \left(\frac{\Delta E_{Ai}^o}{kT}\right) \mathrm{d}\Delta E_{Ai}^o}$$

The noted formula should also apply to conditions for which probabilities are unequal, thus it can be applied as a general expression:

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2161v2 | CC BY 4.0 Open Access | rec: 6 Jul 2016, publ: 6 Jul 2016

$$P_{AB}^{o} = \frac{P_{AB}^{o}}{\sum_{i}^{n} P_{Ai}^{o}}$$

$$= \frac{\int_{a_B}^{b_B} exp - \left(\frac{\Delta E_{AB}^{o}}{kT}\right) \mathrm{d}\Delta E_{AB}^{o}}{\sum_{i}^{n} \int_{a_i}^{b_i} exp - \left(\frac{\Delta E_{Ai}^{o}}{kT}\right) \mathrm{d}\Delta E_{Ai}^{o}}$$

As ΔE_{AB}^{o} increases indefinitely, the probability approaches zero.

$$P_{AB}^{o} = \lim_{\Delta E_{AB}^{o} \to \infty} \frac{\int_{a_B}^{b_B} exp - \left(\frac{\Delta E_{AB}^{o}}{kT}\right) \mathrm{d}\Delta E_{AB}^{o}}{\sum_{i}^{n} \int_{a_i}^{b_i} exp - \left(\frac{\Delta E_{Ai}^{o}}{kT}\right) \mathrm{d}\Delta E_{Ai}^{o}} \cong 0$$

Also, as n increases indefinitely, the probability approaches zero.

$$P_{AB}^{o} = \lim_{n \to \infty} \frac{\int_{a_B}^{b_B} exp - \left(\frac{\Delta E_{AB}^{o}}{kT}\right) \mathrm{d}\Delta E_{AB}^{o}}{\sum_{i}^{n} \int_{a_i}^{b_i} exp - \left(\frac{\Delta E_{Ai}^{o}}{kT}\right) \mathrm{d}\Delta E_{Ai}^{o}} \cong 0$$

We can therefore conclude that, irrespective of condition, the probability, P_{B}^{o}, of finding a bind site that assumes a given conformation B at equilibrium, is approximately equal to zero ($P_{B}^{o} \cong 0$) if CE of conformation B far exceeds that of conformation A; and/or if a substantially large number of conformations occur within the set.

14. If $P_{AB}^{o} \cong 0$ is interpreted as $P_{AB}^{o} \neq 0$, then, based on this interpretation, all possible conformations can occur within the set of conformations. However, it is important to also note that such apparent all-inclusiveness may be relevant for some inconsequential considerations, but for all practical purposes can be deemed irrelevant. Thus, although in principle, the number of members within the set of conformations may increase indefinitely, the number of relevant conformations within a set should remain constant. For example, let us again suppose that at equilibrium, the total number of conformations that can be transitioned from an initial conformation A is n conformations. As previously stated, the sum of all probability values, irrespective of number of conformations must equal unity.

$$\lim_{n \to \infty} \sum_{i}^{n} P_{Ai}^{o} = \cdots + P_{Aw}^{o} + \cdots = 1$$

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2161v2 | CC BY 4.0 Open Access | rec: 6 Jul 2016, publ: 6 Jul 2016

From the above limit, we can express the sum total of probabilities for the first $\boldsymbol{w}$ conformations of the corresponding sequence as:

$$\sum_{i}^{w} P_{Ai}^{o} = \cdots + P_{Aw}^{o}\,, \qquad w < n$$

If

$$\sum_{i}^{w} P_{Ai}^{o} \cong 1$$

Then the sum total of probabilities of conformations within the given set at equilibrium is,

$$\lim_{n\to\infty} \sum_{i}^{n} P_{Ai}^{o} \cong \sum_{i}^{w} P_{Ai}^{o}$$

Whereas $\lim_{n\to\infty} \sum_{i}^{n} P_{Ai}^{o}$ may be considered the sum probabilities of all conformations within the set of conformations, at equilibrium, $\sum_{i}^{w} P_{Ai}^{o}$ may be considered the sum probabilities of relevant conformations within the set at equilibrium; where irrelevant conformations are those with probability values that approach zero. Thus,

$$\lim_{n\to\infty} \sum_{i}^{n} P_{Ai}^{o} = \sum_{i}^{w} P_{Ai}^{o} = 1$$

Thus, we can substitute w for the summation index n in the equation.

$$P_{AB}^{o} = \frac{\int_{a_B}^{b_B} exp - \left(\frac{\Delta E_{AB}^{o}}{kT}\right) \mathrm{d}\Delta E_{AB}^{o}}{\sum_{i}^{w} \int_{a_i}^{b_i} exp - \left(\frac{\Delta E_{Ai}^{o}}{kT}\right) \mathrm{d}\Delta E_{Ai}^{o}}$$

Considering the above relationships, we can therefore express P_{AB}^{o} as a function of ΔE_{AB}^{o} and w.

$$P_{AB}^{o} = p(\Delta E_{AB}^{o}, w)$$

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2161v2 | CC BY 4.0 Open Access | rec: 6 Jul 2016, publ: 6 Jul 2016

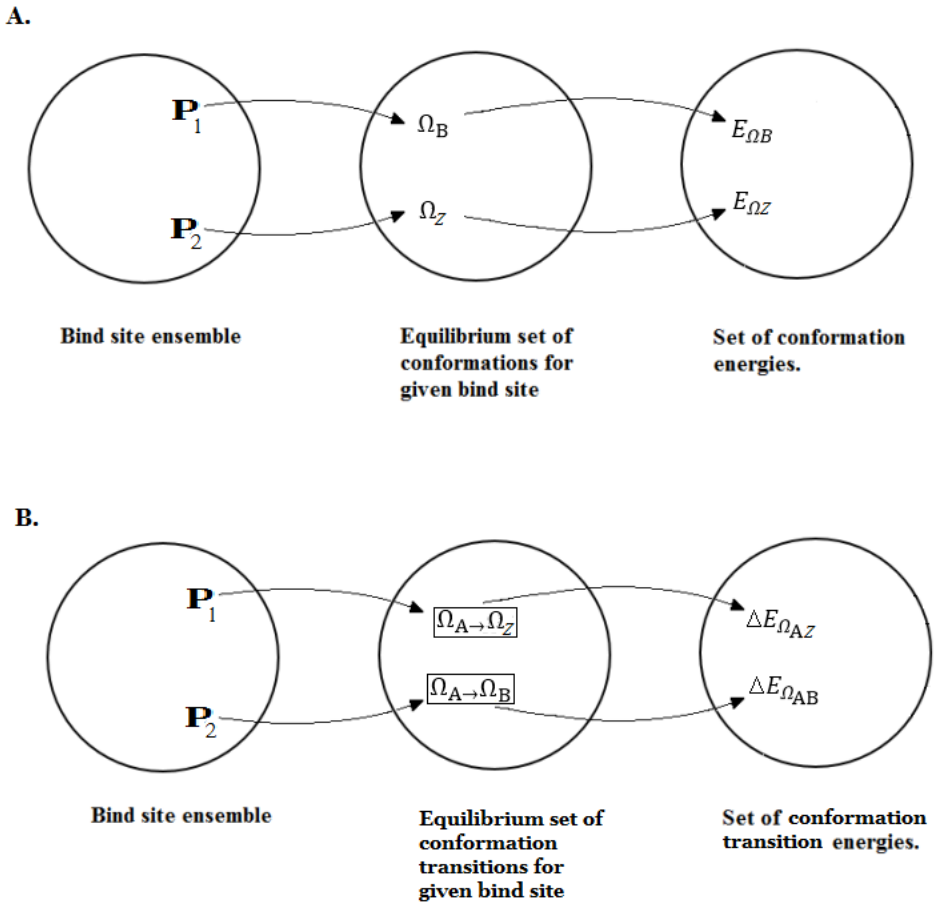

Figure 2A: Relationships between bind site ensemble, equilibrium set of conformations, and set of conformation energies (CE). At any given moment, each bind site of the ensemble maps to a conformation within the equilibrium set of conformations. Two or more bind sites can map to the same conformation, but a given bind site can only map to one conformation. In turn, each conformation maps to a single conformation energy within the set of conformation energies. **Figure 2B**: An illustration of relationships between bind site ensemble, equilibrium set of conformational transitions, and set of conformational transition energies (CTE). Bind sites within ensemble map to conformational transitions within the equilibrium set of conformational transitions. In turn, each conformational transition maps to a single CTE within the set of CTEs.

Probability of finding a given bind site conformation on a physiologically folded peptide under equilibrium conditions and in the presence of ligand.

For the following, we shall apply the term *bind sites* as a general term for description of the non-substrate component of a binding complex. Bind sites as used here refer to both enzyme active site and solvent hydration cage structures. Thus, when applied, the reader should consider such generalization.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2161v2 | CC BY 4.0 Open Access | rec: 6 Jul 2016, publ: 6 Jul 2016

On effects of substrate binding on the bind site structure: reorganization and disorganization of bind site dipole orientations.

15. Here we emphasize the effect of substrate binding to the active site of an enzyme. To understand this, we shall first revisit the relationship between binding and enzyme catalysis event.

16. Natural enzymes are known to catalyze reactions which would otherwise be too slow to support life. The rate accelerations by natural enzymes (as compared to uncatalyzed reactions) can reach $10^6 - 10^{17}$.[9,10] Both experimentation and computer simulations have been employed in the study of enzyme catalyzed reactions. Of particular interest is elucidating the means by which natural enzymes achieve large k_{cat} values. These studies follow from Pauling's proposed role of enzymes in catalysis: In binding and stabilization of the substrate transition state (**TS-substrate**), thereby decreasing the activation energy for conversion from substrate ground state (**GS-substrate**)- to **TS-substrate**[11]. It is this reduction in activation energy that yields an increase in k_{cat}. For these studies, enzyme catalyzed reactions were compared to reactions occurring in aqueous solution and in an absence of enzymes –uncatalyzed reaction.

17. It was posited that a reasonable measure for comparison is the activation energy: where $\Delta g^{\ddagger}_{cat}$ is the activation energy for the enzyme catalyzed reaction; and $\Delta g^{\ddagger}_{cage}$ is the activation energy for the uncatalyzed reaction[12,13]. Also compared were solvation free energies, ΔG_{sol}, of substrates in solvent and active site. The activation energy and solvation free energy are related in that the reduction in activation energy for the enzyme catalyzed reaction is believed to ultimately derive from the solvation free energy, ΔG_{sol}.

18. It was also proposed that the free energy of solvation, ΔG_{sol}, of a cognate TS-substrate when in an active site environment is approximately equal to that for the [bulk] solvent environment. However, unlike within the bulk solvent where the free energy of solvation, ΔG_{sol}, derives, in its entirety, from water molecules interacting (via hydrogen bonds) with TS-substrate; that for an enzyme active site derives from interactions between TS-substrates and both induced and permanent dipoles of residues within the active site. It should also be noted that hydrogen bonding involving water molecules may also contribute to ΔG_{sol} for TS-substrate in an active site environment[14]. However, such contributions pale in comparison to that for the bulk solvent.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2161v2 | CC BY 4.0 Open Access | rec: 6 Jul 2016, publ: 6 Jul 2016

19. The electrostatic nature of these interactions would require that stabilization of TS-substrate involve stabilization of charge(s) on TS-substrate[14]. Maximal stabilization should require optimization of these interactions. In turn, such optimization should require proper alignment of bind site dipoles in relation to charges on TS-substrate. Thus, if initial interactions are non-optimal, then realignment of key dipoles on bind site structure must occur so as to affect [near] optimal interactions[14]. This would require that free energy be invested in the realignment process. The given process of dipole realignment is termed **reorganization**, which we shall properly define later. Thus, the solvation free energy, ΔG_{sol}, can be partitioned into: the free energy of [electrostatic] interactions between TS-substrate and bind site, $\Delta G_{\mu Q}$; and free energy invested in realignment of bind site dipole-dipole interactions, $\Delta G_{\mu\mu}$. As previously stated, such rearrangements facilitate formation of a dipole structure that favors [near] maximum stabilization of TS-substrate[15].

$$\Delta G_{sol} = \Delta G_{\mu Q} + \Delta G_{\mu\mu}$$

20. At this juncture we shall apply designations which are intended to help differentiate enzyme catalyzed reactions from uncatalyzed reactions. It is important to note that the cited works do not use these designations. However, we believe that in order to diminish ambiguities that may arise, these parameters should be specified. Henceforth, the solvation free energy for solvent and active site shall be designated $\Delta G_{sol_{uncat}}$ and $\Delta G_{sol_{cat}}$, respectively. Also, the free energy of [electrostatic] interaction between TS-substrate and solvent molecules shall be designated , $\Delta G_{\mu Q_{uncat}}$, the free energy of [electrostatic] interaction between TS-substrate and active site shall be designated, $\Delta G_{\mu Q_{cat}}$; the free energy invested in rearrangement of solvent dipole-dipole interactions shall be designated, $\Delta G_{\mu\mu_{uncat}}$; and the free energy invested in rearrangement of active site dipole-dipole interactions shall be designated, $\Delta G_{\mu\mu_{cat}}$. Thus, the above equation can be written for the uncatalyzed reaction:

$$\Delta G_{sol_{uncat}} = \Delta G_{\mu Q_{uncat}} + \Delta G_{\mu\mu_{uncat}}$$

And for the catalyzed reaction:

$$\Delta G_{sol_{cat}} = \Delta G_{\mu Q_{cat}} + \Delta G_{\mu\mu_{cat}}$$

In addition, $\Delta\Delta g_{cage}$ is the difference between free energy of activation and the free energy of interactions between the TS-substrate and hydration cage structure.

$$\Delta\Delta g_{cage} = \Delta g^{\ddagger}_{cage} - \Delta G_{\mu Q_{uncat}}$$

$\Delta\Delta g_{cat}$ is the difference in free energy of activation and the free energy of interactions between the substrate transition state and active site.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2161v2 | CC BY 4.0 Open Access | rec: 6 Jul 2016, publ: 6 Jul 2016

$$\Delta\Delta g_{cat} = \Delta g^{\ddagger}_{cat} - \Delta G_{\mu Q}{}_{cat}$$

21. It was proposed that differences between reaction rates of the catalyzed and uncatalyzed reactions are most likely as a result of differences in the energy required to realign dipoles for water molecules of solvent and amino acid residues of active site in relation to the charged TS-substrate[15,17]. In other words, the difference between $\Delta\Delta g_{cat}$ and $\Delta\Delta g_{cage}$ most likely results from differences in $\Delta G_{\mu\mu}$ for enzyme active site and solvent[16]. Rewriting the above equations in terms of $\Delta G_{\mu\mu}$

$$\Delta\Delta g_{cat} = \Delta g^{\ddagger}_{cat} - \Delta G_{sol}{}_{cat} + \Delta G_{\mu\mu}{}_{cat}$$

And

$$\Delta\Delta g_{cage} = \Delta g^{\ddagger}_{cage} - \Delta G_{sol}{}_{uncat} + \Delta G_{\mu\mu}{}_{uncat}$$

Warshel proposed that, $\Delta G_{\mu\mu}$ for uncatalyzed reactions is greater than that for catalyzed reactions[15]. That is,

$$\Delta G_{\mu\mu}{}_{uncat} > \Delta G_{\mu\mu}{}_{cat}$$

The greater $\Delta G_{\mu\mu}$ for uncatalyzed reactions represents a greater amount of free energy invested in formation of the hydration cage structure[16]. Recall that $\Delta G_{\mu\mu}$ reflects the energy invested in the process termed reorganization.

22. **Reorganization** involves changes in bind site dipole-dipole interactions and are required to achieve [near] optimal alignment of involved residues, in the case of active site, or water molecules in the case of formation of hydration cage. As previously stated, such realignment results in a dipole organization that [almost] maximally stabilizes TS-substrate[15]. A reasonable inference that can be drawn from these proposals and findings is that the greater $\Delta G_{\mu\mu}$ for reorganization of a hydration cage, $\Delta G_{\mu\mu}{}_{uncat}$, as compared to that for reorganization of the active site, $\Delta G_{\mu\mu}{}_{cat}$, must derive from the greater extent of disorganization of the hydration cage as compared to the active site structure.

Here **disorganization** refers to the extent to which the initial dipole-dipole orientations differ from the fully reorganized orientations. With dipole alignments that deviate from the maximally stabilizing orientations having a greater extent of disorganization than those which show less

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2161v2 | CC BY 4.0 Open Access | rec: 6 Jul 2016, publ: 6 Jul 2016

deviation. The noted inference is motivated by the position that enzyme active sites are believed to have a preorganized structure that already affects partial stabilization of TS-substrate, and thus requires little reorganization. Thus, explaining the lower values of $\Delta G_{\mu\mu}$ for enzyme-based– than for solvent-based- reactions[15]. In other words, the active site structure undergoes less disorganization than the bulk solvent.

Structural reorganization and disorganization

23. In addition to induced dipoles, permanent dipoles (e.g., charges on active site residues), have been proposed to also be involved in stabilization of substrate transition state[14]. An example of transition state stabilization by permanent dipoles (hydrogen bonds) is the means by which the hydration cage structure stabilizes the charged transition state substrate when in an absence of catalysts. For these uncatalyzed reactions, the optimally stabilizing bind site, solvent cage structure, must be organized either from an initial poorly stabilizing structure[18] or an amorphous, and thus highly disorganized structure (figure 3B). The noted process can be considered the structural component of reorganization for the solvent-based reaction, whereas the reverse is considered the structural component of disorganization for the solvent-based reaction (figure 3B). Such considerations are based on the notion that, in principle, alignment of these [permanent] dipoles may involve structural changes as the aligning units assume orientations that approximate optimal interactions with TS-substrates.

24. Two relevant questions are of interest: 1) whether proper alignment of permanent dipoles in relation to substrate transition state require changes to active site structure? and 2) what is the lower limit for the extent of structural change that can be considered significant.

Regarding the first question, consider that for dipole-dipole reorganization events involving induced dipoles mere electron dispersal may occur for a fixed active site structure. However, for reorganization events involving permanent dipoles –that is to alter the orientation of the given permanent dipole relative to the substrate– bond rotations and thus changes to molecular configurations must occur. Again, consider the changes in orientation of water molecules that must occur in order to form the hydration cage.

In principle, the extent to which active site structures are altered can range from such small single bond rotations –with changes in side chain dihedral angles– to more widespread and thus profound changes. Of these changes, it is a reasonable stance to consider that widespread changes have a significant impact on catalytic rates. Based on the same line of reasoning, one may consider that diminishingly smaller alterations would yield negligible change in catalytic rates. Thus, regarding the second question, the lower limit of structural

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2161v2 | CC BY 4.0 Open Access | rec: 6 Jul 2016, publ: 6 Jul 2016

changes may be erroneously considered to involve extensive displacement of involved segment. However, significant alterations in catalytic rates have been shown to be associated with insignificant changes in active site structure. For example, Mesecar and Koshland (1997) showed that affecting isocitrate dehydrogenase structure such that the resultant changes in amino acid positions are less than an Angstrom can result in significant orders of magnitude change in catalytic rates[19].

25. Thus, in principle, the reorganization event may also comprise significant conformational (structural) changes. We consider such conformational changes as the structural component of reorganization for active site. We also consider the extent of disorganization of these permanent dipoles, and thus bind site structures, prior to reorganization as a determinant of the degree of change in both structural features that must occur to yield the given transition state stabilizing structure. This can be considered the structural component of disorganization for active site. Also, a structural component of preorganization would exist for the active site (figure 4), and should explain the difference between the extent of structural reorganization that must occur for active site and bulk solvent.

26. Based on such explicit distinction between induced and permanent dipole contributions, it should hold that a fraction of the free energy of reorganization, $\Delta G_{\mu\mu_{cat}}$, may be invested in such structural changes.

$$\Delta G_{\mu\mu_{cat}} = x\Delta G_{\mu\mu_{cat}} + (1-x)\Delta G_{\mu\mu_{cat}}$$

Where,

$x\Delta G_{\mu\mu_{cat}}$ = Fraction of $\Delta G_{\mu\mu_{cat}}$ that affects structural reorganization

$(1-x)\Delta G_{\mu\mu_{cat}}$ = Fraction of $\Delta G_{\mu\mu_{cat}}$ that affects nonstructural reorganization

Structural reorganization and ligand-associated probability bias.

27. Suppose an initial condition of which all bind sites of the ensemble are saturated with substrates and that the substrate is at just the threshold concentration to yield such saturated bind sites.

The probability of finding a bind site that assumes a given conformational transition involving transition from an initial conformation ***A*** and transition into conformation ***B*** increases with either: increasing conformational energy (CE) of conformation ***A*** for a fixed CE of conformation ***B*** (figure 5A); decreasing CE of conformation ***B*** for a fixed CE of conformation ***A*** (figure 5B); or both a reduction of CE for conformation ***B*** and an increase in CE of conformation ***A*** (figure 5C).

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2161v2 | CC BY 4.0 Open Access | rec: 6 Jul 2016, publ: 6 Jul 2016

The probability of finding a bind site that assumes a given conformational transition involving transition from an initial conformation ***A*** and transition into conformation ***Z*** decreases with either: an increase in CE of conformation ***Z*** for a fixed CE of conformation ***A*** (figure 6A); or reduction of CE of conformation ***A*** for a fixed CE of conformation ***Z*** (figure 6B), or both an increase in CE for conformation ***Z*** and a reduction in CE of conformation ***A*** (figure 6C).

In essence, the probability of finding a bind site that undergoes conformational transition from conformation ***A*** to conformation ***B*** increases with reduction in CTE for transition from conformation ***A*** to conformation ***B***. Also, the probability of finding a bind site that undergoes conformation transition from conformation ***A*** to conformation ***Z*** decreases with an increase in CTE for transition from conformation ***A*** to conformation ***Z***.

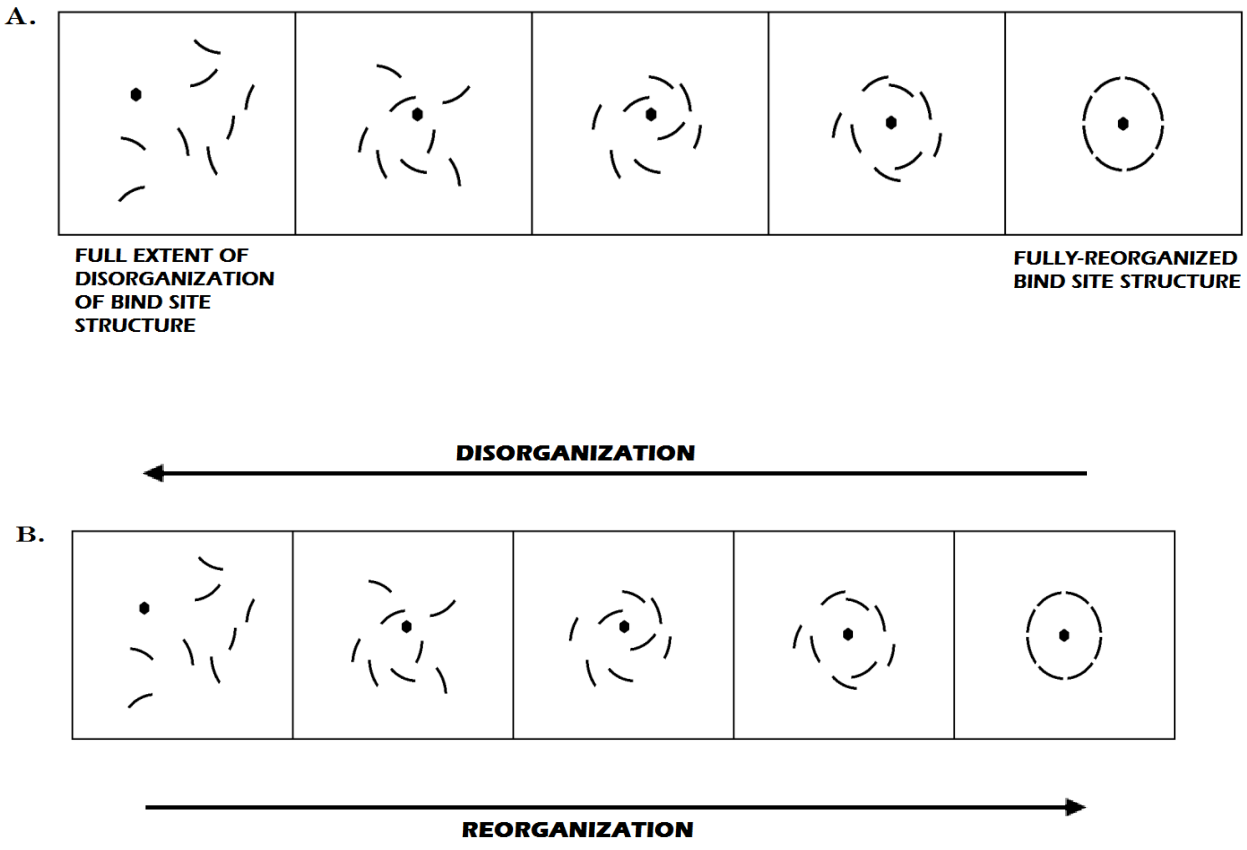

Figure 3A: Depicts the range of structures of a bind site. At one extreme is a much disorganized structure termed the full-extent of disorganization of bind site structure. At the other extreme is

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2161v2 | CC BY 4.0 Open Access | rec: 6 Jul 2016, publ: 6 Jul 2016

the fully-reorganized bind site structure. **Figure 3B**: Illustrates the direction of reorganization and disorganization. For reorganization, the bind site structure is organized from the full-extent of the disorganized structure to the fully-reorganized bind site structure. Disorganization involves a reversal of the noted direction for reorganization. That is, from the fully-reorganized bind site structure, the bind site is transformed to the full-extent of the disorganized structure.

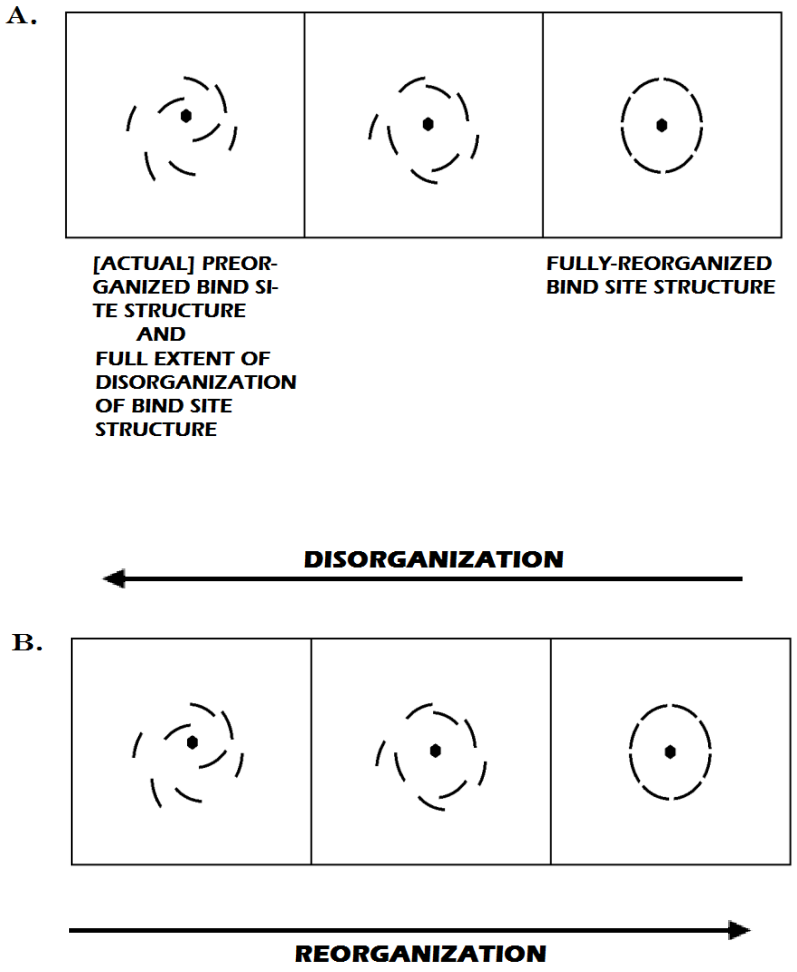

Figure 4A: Depicts the range of structures of a preorganized bind site. Compare the full-extent of disorganization of bind site structure to that of figure 3. Note that the full extent of disorganization of bind site is more organized than that of figure 3. In other words, the most disorganized structure has been preorganized so as to prevent the extent of disorganization noted for figure 3. **Figure 4B**: Illustrates the direction of reorganization and disorganization. For reorganization, the bind site structure is organized from a preorganized structure to the fully-

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2161v2 | CC BY 4.0 Open Access | rec: 6 Jul 2016, publ: 6 Jul 2016

reorganized bind site structure. Disorganization involves a reversal of the noted direction for reorganization. That is from the reorganized bind site structure to the preorganized structure.

If significant increase in CE of conformation $\boldsymbol{Z}$ such that the probability of finding a bind site that assumes conformation $\boldsymbol{Z}$ when at equilibrium approaches zero, then the total number of conformations within set – the upper limit for summation index of probabilities– can therefore be considered to be decreased. Thus, an increase in CE of conformation Z is also proportional to an increase in the probability of finding a bind site that undergoes conformation transition from conformation $\boldsymbol{A}$ to conformation $\boldsymbol{B}$.

A combination of these effects can affect a probability bias mostly in favor of a given conformation or number of conformations. Thus, some conformational probabilities may require that some CEs be increased, whereas others are decreased or stay unchanged

28. Let us now suppose that the reorganized structure is defined by conformation B and only by conformation B. In other words, the reorganized structure occurs if and only if conformation B is assumed. In addition, we suppose that conformation B can be said to be assumed if and only if the reorganized structure occurs. Thus, conformation transition from conformation A to conformation B can be considered a structural reorganization event. Thus, in a similar way as formation of appropriate [electrostatic] interactions between TS-substrate and active site environment can affect a reduction in activation energy for the given change in substrate state, so too can these same interactions affect reduction in conformational transition energy as the actives site undergoes reorganization. We define the effective CTE when cognate ligand is present and bind site is saturated with cognate ligand as: as the difference between ΔE^{o}_{AB} and $x\Delta G_{\mu\mu_{cat}}$ for the peptide molecule.

$$\Delta E_{AB} = \Delta E^{o}_{AB} + x\Delta G_{\mu\mu_{cat}}$$

Thus in addition to decreasing $\Delta \mathrm{g}^{\ddagger}_{cat}$ for conversion from GS-substrate to TS-substrate, such binding interactions may also decrease the conformational transition energy required for transition from conformation A to conformation B.

We can determine the probability, P_{AB}, of conformation transition involving transition from an initial conformation $\boldsymbol{A}$, within the set of conformations, and transition into conformation $\boldsymbol{B}$ as opposed to a conformation $\boldsymbol{Z}$ within the set of conformations.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2161v2 | CC BY 4.0 Open Access | rec: 6 Jul 2016, publ: 6 Jul 2016

$$P_{AB} = \frac{P_{AB}}{\sum_i^n P_{Ai}} = \frac{\int_{a_B}^{b_B} exp - \left(\frac{\Delta E_{AB}}{kT}\right) \mathrm{d}\Delta E_{AB}}{\sum_i^n \int_{a_i}^{b_i} exp - \left(\frac{\Delta E_{AB}}{kT}\right) \mathrm{d}\Delta E_{Ai}}$$

Considering the above relationships, we can therefore express P_{AB} as a function of ΔE_{AB} and w

$$P_{AB} = p(\Delta E_{AB}, w)$$

29. We define a measure of conformational transition probability bias as the extent of ligand effect on the probability of finding a bind site, within the ensemble, that assumes a given conformational transition.

$$Probability\ bias = P_{AB} - P_{AB}^{o}$$

If $P_{AB} - P_{AB}^{o} > 0$, then a higher probability of finding a bind site that assumes the given conformational transition when in the presence of- than in an absence of ligand. If $P_{AB} - P_{AB}^{o} < 0$, then a lower probability of finding a bind site that assumes the given conformation transition when in the presence of- than in an absence of ligand. If $P_{AB} - P_{AB}^{o} = 0$, then there is no difference in probability of finding a bind site that assumes the given conformational transition when in the presence of- than in an absence of ligand.

A view of reorganized and rigidified bind site structures as [apparent] preorganized structures.

30. As stated previously, during formation of the active site: substrate complex, the active site can be viewed as initiating from a less organized structure –limits of which we consider the full extent of disorganization. It evolves to a more organized structure–limits of which we consider the fully reorganized structure. A reversal of this process should also occur for loss of complex (figures). Also, CT cycles of active site are between those conformations that define the full extent of disorganization for active site and those that define fully-reorganized structures.

31. As compared to the bulk solvent, the active site undergoes less disorganization (compare figures 3 and 4). As was previously stated, this is due to the preorganized structure of the active

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2161v2 | CC BY 4.0 Open Access | rec: 6 Jul 2016, publ: 6 Jul 2016

site. Warshel posited that formation of the preorganized structure occurs during the folding process[13]. Henceforth, we shall designate the preorganized structure that derives from this early preorganization event as an [actual] preorganized structure. Also, the [actual] preorganized structure is maintained throughout the life-time of the active site when in native state. As compared to active sites, water molecules of the solvent (uncatalyzed reaction) cannot maintain such preorganized structures. Hence the greater extent of disorganization than their active site counterparts.

This ability to maintain a preorganized structure stems from differences in rigidity of the structures involved[17]. It has been proposed that such rigidity may derive, in part, from the large sizes of enzymes, as gauged by the positive correlation between peptide size and heat capacity[10]. In addition, the entirety of intra-peptide interactions may favor significant rigidity. Other sources of rigidity may be extra-peptide interactions, as may occur between peptide and milieu constituents. Of these extra-peptide interactions, those involving cognate ligand are especially significant for this work. There is evidence from literature that binding of cognate ligand can increase the rigidity of peptide structures[20-22].

32. Thus, we deduce from the above analysis that, at least ***in theory***, reorganization events that are followed by *permanent and significant increments in rigidity of the fully reorganized structure should yield an [apparent] preorganization*. The "apparent" term is intended to differentiate such seemingly preorganized structures from those that result from "actual" preorganization that occurs during folding. Thus, in an absence of prior knowledge, this initial structure may be considered to have derived from [actual] preorganization.

A.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2161v2 | CC BY 4.0 Open Access | rec: 6 Jul 2016, publ: 6 Jul 2016

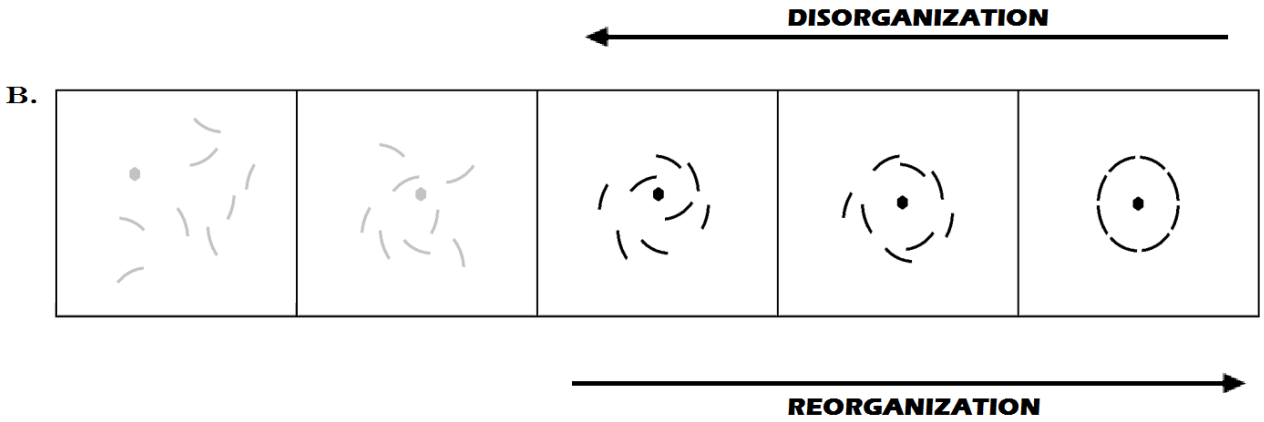

Figure 7A: Depicts the range of structures of a bind site. At one extreme is the full-extent of disorganization of bind site structure. However, note that although the full-extent of disorganized structure is identical to that for figure 4, the means by which they derive differ. Whereas [actual] preorganization results in that noted for figure 4, [apparent] preorganization results for that noted for figure 7. Figure 7B: Illustrates the direction of reorganization and disorganization. Reorganization from an [apparent] preorganized structure, as compared to that occurring from full-extent of disorganized structures as in figure 3, would require less free energy of reorganization.

A view of [actual] preorganized structures as resulting from reorganization events that occur during peptide folding.

33. As previously stated, formation of the [actual] preorganized structure occurs during the folding process. That is, an enzyme that starts out in an unfolded state, for which little to none of its native state features can be appreciated must undergo alignment and realignment of its constituent amino acid residues until the final native product results. In other words, the active site can be viewed as evolving from a less organized- to a more organized structure; where organization reflects an increasing resemblance to an active site structure when complexed with substrate. Thus, formation of [actual] preorganized structure can be considered analogous to a reorganization event required to yield the fully-reorganized structure. In effect, we may consider the [actual] preorganization event as an incomplete reorganization event that occur during the folding process; with a second or main component of reorganization event occurring post-

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2161v2 | CC BY 4.0 Open Access | rec: 6 Jul 2016, publ: 6 Jul 2016

folding – during formation of the ligand:bind site complex.

34. Thus, we not only have reconciled reorganization with the process of formation of [apparent] preorganized structures, we also reconcile reorganization with the process of formation of [actual] preorganized structures.

35. Thus, ***at least in theory***, just as we considered that *reorganization events that are followed by permanent and significant increments in rigidity of the reorganized structure* would yield an [apparent] preorganized structure, this same approach should yield an [actual] preorganized structure, if occurring during the folding process.

36. Structures that typify those of unfolded and/or partially folded intermediates of peptides undergoing either folding or refolding have been noted in the literature: $p21^{Wafl/Cip1/Sdi1}$;[23] poplar apo-plastocyanin[24]; Substrate-binding peptide (peptide I) from DNA polymerase I[25]; and a host of others. A list of these peptides can be found in references [26] and [27]. Upon binding its cognate ligand, these so-called "intrinsically disordered peptides" have been shown to yield structures with greater extent of order following binding to their cognate ligands[28].

37. Using this same approach and a similar role of cognate ligands we can attempt at affecting the "organization" of unfolded and/or partially folded intermediates of peptides undergoing folding. That is, upon formation of complex with ligand, the folding peptides should assume more organized structures that are reminiscent of their native fold state counterparts.

38. On this basis, a similar approach to affecting probability bias in favor of conformations that define reorganized structures – and thus in favor of reorganization events – for the folded peptide, should yield similar outcomes if utilized for particular fold states of a folding peptide. That is, presentation of the cognate substrate at a given concentration range should affect reorganization events of bind sites on peptides undergoing folding.

Key considerations when attempting verification of the stated schema

39. To achieve such an early reorganization event, we can apply the approach proposed for formation of [apparent] preorganized bind sites. That is, reorganization using cognate ligands. Since an increase in rigidity typically occur as the peptide transitions from more extended to more compact states, we shall not emphasize any additional attempts at increasing the rigidity of

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2161v2 | CC BY 4.0 Open Access | rec: 6 Jul 2016, publ: 6 Jul 2016

the peptide molecule. In other words, rigidity that occurs with the folding process is assumed to be sufficient.

40. As was previously stated enzyme active sites undergo [actual] preorganization events and are therefore preorganized. Hence the noted difference between the extent of reorganization for an active site and that for a hydration cage structure (compare figure 3 and 4). Since the [actual] preorganization event is furnished solely by the free energy of folding[16,29], the noted event need not be facilitated by cognate ligand, and therefore should occur in an absence of the cognate ligand. Thus, one cannot conclude with a high degree of certainty that an [actual] preorganization event that may occur in the presence of ligand is facilitated by the cognate ligand. A solution to this problem is to consider that even for an [actual] preorganized bind site that is furnished by the free energy of folding, the given bind site must still undergo reorganization to a fully-reorganized bind site structure, albeit to a lesser extent than the hydration cage structure (compare figure 3 and 4); thus, $\Delta G_{\mu\mu_{cat}} \neq 0$. Based on the proposed role of preorganization in diminishing the extent of reorganization required to yield a fully reorganized structure, we posit that: [actual] preorganized bind site structures that are identical to the fully-reorganized structure should not require reorganization (figure 8); that is, $\Delta G_{\mu\mu_{cat}} \cong 0$. Thus, if in the absence of cognate ligand, $\Delta G_{\mu\mu_{cat}} \neq 0$, and in the presence of cognate ligand, $\Delta G_{\mu\mu_{cat}} \cong 0$, then for such cases, we can determine with a high degree of certainty that [actual] preorganization events with $\Delta G_{\mu\mu_{cat}} \cong 0$ that may occur in the presence of ligand is influenced by the cognate ligand.

41. In the same way we defined parameters for comparison of catalyzed to uncatalyzed reactions (reference), we can define similar parameters for comparison of natural enzymes undergoing folding in the presence of cognate ligand to those undergoing folding in an absence of ligand (reference). A reasonable measure is the probability, at equilibrium, of finding a bind site within the ensemble that assumes a given conformation which defines the fully-reorganized bind site structure.

42. However, different fold states of the peptide may yield different probability values. For example, the probability, at equilibrium, of finding bind sites within the ensemble that assume a given conformation which define the fully-reorganized bind site structure is greater for the population of enzymes in native fold state than for those in a fully extended state. As a consequence, the probability bias in favor or against conformation(s) that define the fully-reorganized bind site may also vary for different fold states. Reasonable measures are **probability maxima** for each probability measure (in the presence of- versus in an absence of cognate ligand); and the difference between these maxima as was done in determination of probability bias. Fold state, as used here, refer to the extent to which the peptide is folded, with the possible fold states being: unfolded (fully-extended); native (physiologically folded); or intermediate(s)

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2161v2 | CC BY 4.0 Open Access | rec: 6 Jul 2016, publ: 6 Jul 2016

(partially folded).

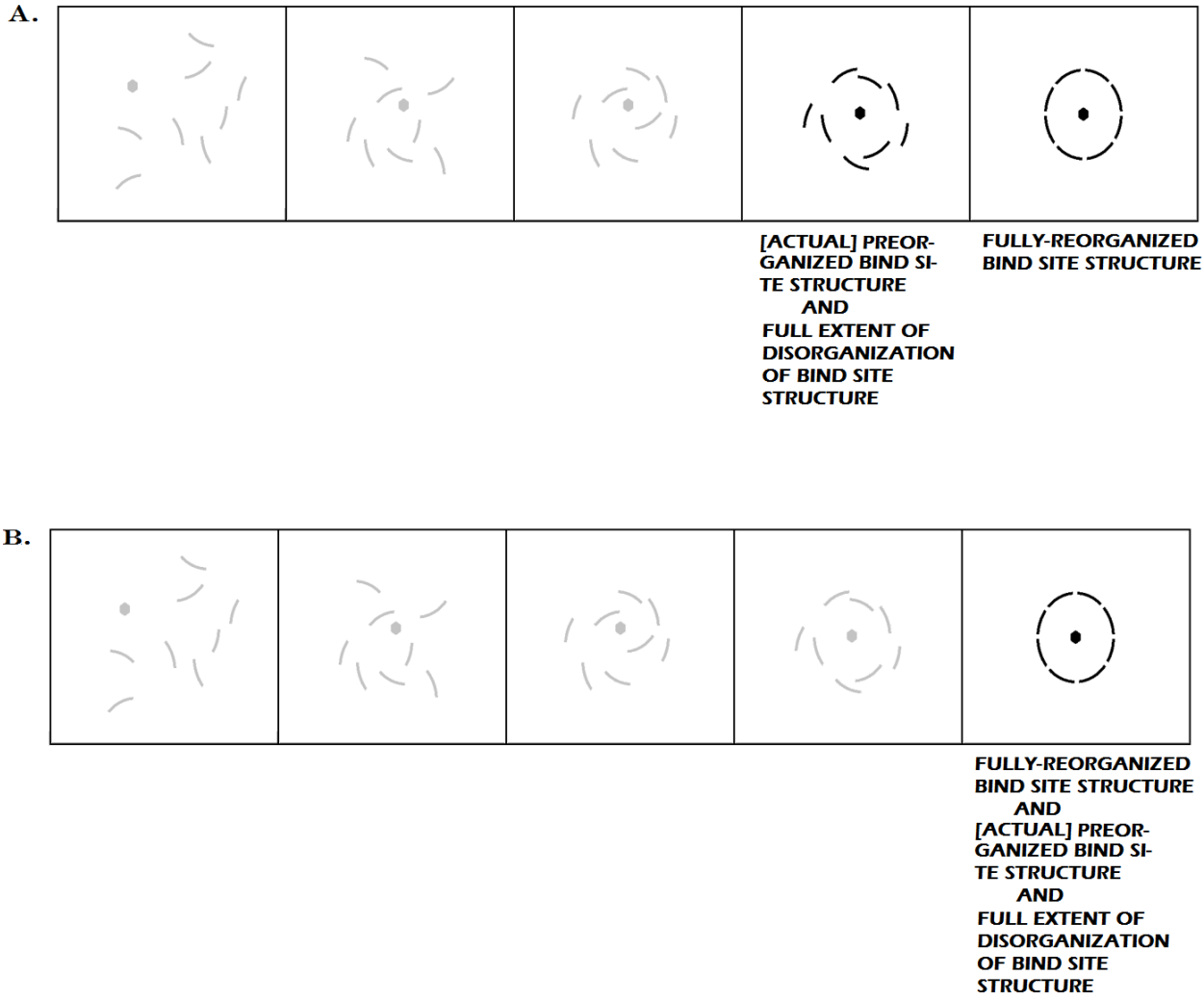

Figure 8.

Figure 8A: The preorganized structure is much ordered and therefore undergoes less disorganization than that for figure 4 and 7. **Figure 8B**: The preorganized structure is the same as the fully-reorganized structure and would therefore not undergo disorganization. Compare figures 8 to figures 7.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2161v2 | CC BY 4.0 Open Access | rec: 6 Jul 2016, publ: 6 Jul 2016

Peptide fold states at equilibrium

43. We suppose an initial fully unfolded (extended) state of all peptide molecules of ensemble. We now suppose all peptide molecules undergo folding. Suppose a natural enzyme as the peptide of interest, and that folding occurs under conditions that yield a population of native fold state; then at equilibrium there should exist at least two populations of peptide fold states: A population of peptides assuming a fully extended (unfolded) fold state and a second population assuming a physiologically-folded (native) fold state. Let us define an equilibrium for the unfolded and native states of a peptide molecule as:

$$[\mathrm{U}] \leftrightarrow [\mathrm{N}]$$

44. In principle, additional fold state populations may exist, with the fold state being an intermediate of unfolded and native fold states. In other words, these are partially folded intermediates. Let us define the equilibrium for the unfolded, partially folded intermediate, and native fold states of a peptide molecule as:

$$[\mathrm{U}] \leftrightarrow [\mathrm{I}] \leftrightarrow [\mathrm{N}]$$

Thus for an equilibrium involving j populations of partially folded intermediate states:

$$[\mathrm{U}] \leftrightarrow \cdots \leftrightarrow \left[\mathrm{I}_j\right] \leftrightarrow \cdots \leftrightarrow [\mathrm{N}]$$

We shall apply a non-specific symbolic representation, $\boldsymbol{I}_{\Omega}$, for fold states;

Probability of finding a given bind site conformation within the set of bind site conformations in the presence and absence of ligand (TS-substrate):

45. Let us suppose that the *fully* reorganized structure is defined by conformation B and only by conformation B. In other words, the reorganized structure occurs if and only if conformation B is assumed. Let us also suppose that conformation B can be said to be assumed if and only if the reorganized structure occurs. Thus, in order for an [actual] preorganized bind site to be identical to the fully-reorganized bind site structure, it must also be defined by conformation B.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2161v2 | CC BY 4.0 Open Access | rec: 6 Jul 2016, publ: 6 Jul 2016

In addition, we suppose that bind sites of all fold states are saturated with ligand. It was proposed that formation of the preorganized active site structure is furnished by the folding energy, ΔG_{fold} [16, 29]. With the given relationship:

$$\Delta G_{\mu\mu} = \Delta G_{fold}$$

In terms of conventions used here, we can rewrite the above equation as:

$$\Delta G_{\mu\mu_{fold}} = \Delta G_{fold}$$

where, $\Delta G_{\mu\mu_{fold}}$, is the reorganization energy of the bind site during folding.

Thus, as a starting point, we define the CTE for transition from an initial conformation ***A***, within the set of conformations, and transition into conformation *B* for the given fold state, $\boldsymbol{I}_{\Omega}$, of a peptide undergoing folding and in an absence of cognate ligand, $\Delta E^{o}_{AB_{fold}}$, as*:*

$$\boldsymbol{I}_{\Omega}(\Delta E^{o}_{AB_{fold}}) = \boldsymbol{I}_{\Omega}(\Delta \mathbf{e}_{AB} + \Delta G_{fold})$$

where $\Delta \mathbf{e}_{AB}$ is the CTE for transition from conformation A to conformation B for the given fold state of a peptide not undergoing folding and in an absence of cognate ligand; ΔG_{fold} is the free energy of peptide folding.

Similarly, we define the CTE for transition from an initial conformation ***A***, within the set of conformations, and transition into conformation *B* for the given fold state, $\boldsymbol{I}_{\Omega}$, of a peptide undergoing global folding and in the **presence** of cognate ligand, $\Delta E^{o}_{AB_{fold}}$, as*:*

$$\boldsymbol{I}_{\Omega}(\Delta E_{AB_{fold}}) = \boldsymbol{I}_{\Omega}(\Delta E^{o}_{AB_{fold}} + x\Delta G_{\mu\mu_{fold}})$$

where $x\Delta G_{\mu\mu_{fold}}$ is the fraction of $\Delta G_{\mu\mu_{fold}}$ that affects structural reorganization of the peptide at given fold state and can be expressed as:

$$x\Delta G_{\mu\mu_{fold}} = \Delta G_{sol_{fold}} - \Delta G_{\mu Q_{fold}} - (1-x)\Delta G_{\mu\mu_{fold}}$$

Substituting for $x\Delta G_{\mu\mu_{fold}}$

$$\boldsymbol{I}_{\Omega}(\Delta E_{AB_{fold}}) = \boldsymbol{I}_{\Omega}(\Delta E^{o}_{AB_{fold}} + \Delta G_{sol_{fold}} - \Delta G_{\mu Q_{fold}} - (1-x)\Delta G_{\mu\mu_{fold}})$$

Substituting for $\Delta E^{o}_{AB_{fold}}$

$$\boldsymbol{I}_{\Omega}(\Delta E_{AB_{fold}}) = \boldsymbol{I}_{\Omega}(\Delta \mathbf{e}_{AB} + \Delta G_{fold} + \Delta G_{sol_{fold}} - \Delta G_{\mu Q_{fold}} - (1-x)\Delta G_{\mu\mu_{fold}})$$

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2161v2 | CC BY 4.0 Open Access | rec: 6 Jul 2016, publ: 6 Jul 2016

Where, $\Delta G_{sol_{fold}}$, is the solvation free energy for TS substrate when interacting with bind site on peptide at given fold state; $\Delta G_{\mu Q_{fold}}$, is the free energy of interaction between the TS substrate and peptide at given fold state; $(1-x)\Delta G_{\mu\mu_{fold}}$, is the fraction of $\Delta G_{\mu\mu_{fold}}$ that affects non-structural reorganization.

From the view point of preorganization as an early reorganization event, it should follow that folding energy[16,29] would be considered a source of free energy of reorganization. Specifically, the free energy of structural reorganization is:

$$x\Delta G_{\mu\mu_{fold}} = \Delta G_{fold} + \Delta G_{sol_{fold}} - \Delta G_{\mu Q_{fold}} - (1-x)\Delta G_{\mu\mu_{fold}}$$

In this sense, the given equality:

$$\Delta G_{\mu\mu_{fold}} = \Delta G_{fold}$$

Can be considered a special condition for which $\Delta G_{sol_{fold}} - \Delta G_{\mu Q_{fold}} = 0$. In other words, when the cognate ligand is absent.

A proper representation of $\Delta E_{AB_{fold}}$ for a given fold state is therefore:

$$\begin{aligned} \boldsymbol{I}_{\Omega}\left(\Delta E_{AB_{fold}}\right) &= \boldsymbol{I}_{\Omega}\left(\Delta \mathbf{e}_{AB} + x\Delta G_{\mu\mu_{fold}}\right) \\ &= \boldsymbol{I}_{\Omega}(\Delta E^{0}_{AB_{fold}} + x\Delta G_{\mu\mu_{fold}} - \Delta G_{fold}) \end{aligned}$$

At equilibrium, we can determine, in the presence of cognate ligand and when bind site is saturated with cognate ligand, the probability, $P_{AB_{fold}}$, of finding a bind site within the ensemble that assumes the given conformation transition involving transition from an initial conformation $\boldsymbol{A}$, within the set of conformations, and transition into conformation $\boldsymbol{B}$ as opposed to a conformation $\boldsymbol{Z}$ within the set of conformations; as:

$$\boldsymbol{I}_{\Omega}(P_{AB_{fold}}) = \boldsymbol{I}_{\Omega}\frac{\int_{a_B}^{b_B} exp - \left(\frac{\Delta E_{AB_{fold}}}{kT}\right) \mathrm{d}\Delta E_{AB_{fold}}}{\sum_{i}^{n}\int_{a_i}^{b_i} exp - \left(\frac{\Delta E_{Ai_{fold}}}{kT}\right) \mathrm{d}\Delta E_{Ai_{fold}}}$$

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2161v2 | CC BY 4.0 Open Access | rec: 6 Jul 2016, publ: 6 Jul 2016

Similarly, at equilibrium, we can determine the probability, $P^{o}_{AB_{fold}}$, of finding a bind site within the ensemble that assumes the given conformation transition involving transition from an initial conformation $\boldsymbol{A}$, within the set of conformations, and transition into conformation $\boldsymbol{B}$ as opposed to a conformation $\boldsymbol{Z}$ within the set of conformations; and in an absence of cognate ligand as:

$$\boldsymbol{I}_{\Omega}(P^{o}_{AB_{fold}}) = \boldsymbol{I}_{\Omega}\frac{\int_{a_B}^{b_B} exp - \left(\frac{\Delta E^{o}_{AB_{fold}}}{kT}\right) \mathrm{d}\Delta E^{o}_{AB_{fold}}}{\sum_{i}^{n}\int_{a_i}^{b_i} exp - \left(\frac{\Delta E^{o}_{Ai_{fold}}}{kT}\right) \mathrm{d}\Delta E^{o}_{Ai_{fold}}}$$

Maximum probability of finding a given bind site conformation within the set of bind site conformations in the presence and absence of ligand (TS substrate):

We can determine the difference between the maximum probability for peptide folding in the presence of cognate ligand, $\boldsymbol{I}_{\Omega}\left(P_{AB_{fold}}\right)_{max}$, and the maximum probability for peptide folding in an absence of ligand, $\boldsymbol{I}_{\Omega}(P^{o}_{AB_{fold}})_{max}$, as:

$$\boldsymbol{I}_{\Omega}\left(P_{AB_{fold}}\right)_{max} - \boldsymbol{I}_{\Omega}(P^{o}_{AB_{fold}})_{max}$$

If,

$$\boldsymbol{I}_{\Omega}\left(P_{AB_{fold}}\right)_{max} - \boldsymbol{I}_{\Omega}(P^{o}_{AB_{fold}})_{max} > 0$$

then a higher maximum probability of finding a bind site that assumes the given conformation B when in the presence of- than when in an absence of ligand.

If,

$$\boldsymbol{I}_{\Omega}\left(P_{AB_{fold}}\right)_{max} - \boldsymbol{I}_{\Omega}(P^{o}_{AB_{fold}})_{max} < 0$$

then a lower maximum probability of finding a bind site that assumes the given conformation B when in the presence of- than when in an absence of ligand.

If,

$$\boldsymbol{I}_{\Omega}\left(P_{AB_{fold}}\right)_{max} - \boldsymbol{I}_{\Omega}\left(P^{o}_{AB_{fold}}\right)_{max} = 0$$

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2161v2 | CC BY 4.0 Open Access | rec: 6 Jul 2016, publ: 6 Jul 2016

then there is no difference in maximum probability of finding a bind site that assumes the given conformation B when in the presence of- than when in an absence of ligand.

References:

[1] Bogan, A. A., & Thorn, K. S. (1998). Anatomy of hot spots in protein interfaces. Journal of molecular biology, 280(1), 1-9.

[2] Fischer, E. (1894). Einfluss der Configuration auf die Wirkung der Enzyme. Berichte der deutschen chemischen Gesellschaft, 27(3), 2985-2993.

[3] Koshland, D. E. (1998). Conformational changes: how small is big enough?. Nature medicine, 4(10), 1112-1114.

[4] Changeux, J. P., & Edelstein, S. (2011). Conformational selection or induced fit? 50 years of debate resolved. F1000 Biol Rep, 3, 19.

[5] Changeux, J. P. (2013). 50 years of allosteric interactions: the twists and turns of the models. Nature reviews Molecular cell biology, 14(12), 819-829.

[6] Gianni, S., Dogan, J., & Jemth, P. (2014). Distinguishing induced fit from conformational selection. Biophysical chemistry, 189, 33-39.

[7] Koshland Jr, D. E., Nemethy, G., & Filmer, D. (1966). Comparison of experimental binding data and theoretical models in proteins containing subunits*. Biochemistry, 5(1), 365-385.

[8] Boehr, D. D., Nussinov, R., & Wright, P. E. (2009). The role of dynamic conformational ensembles in biomolecular recognition. Nature chemical biology, 5(11), 789-796.

[9] Hilvert, D. (2000). Critical analysis of antibody catalysis. Annual review of biochemistry, 69(1), 751-793.

[10] Cannon, W. R., & Benkovic, S. J. (1998). Solvation, reorganization energy, and biological catalysis. Journal of Biological Chemistry, 273(41), 26257-26260.

[11] Pauling, L. (1946). Molecular architecture and biological reactions. Chemical and engineering news, 24(10), 1375-1377.

[12] Villa, J., Štrajbl, M., Glennon, T. M., Sham, Y. Y., Chu, Z. T., & Warshel, A. (2000). How important are entropic contributions to enzyme catalysis? Proceedings of the National Academy of Sciences, 97(22), 11899-11904.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2161v2 | CC BY 4.0 Open Access | rec: 6 Jul 2016, publ: 6 Jul 2016

[13] Warshel, A., Sharma, P. K., Kato, M., Xiang, Y., Liu, H., & Olsson, M. H. (2006). Electrostatic basis for enzyme catalysis. Chemical reviews, 106(8), 3210-3235

[14] Warshel, A. (1978). Energetics of enzyme catalysis. Proceedings of the National Academy of Sciences, 75(11), 5250-5254.

[15] Villa, J., & Warshel, A. (2001). Energetics and dynamics of enzymatic reactions. The Journal of Physical Chemistry B, 105(33), 7887-7907.

[16] Warshel, A. (1998). Electrostatic origin of the catalytic power of enzymes and the role of preorganized active sites. Journal of Biological Chemistry, 273(42), 27035-27038.

[17] Åqvist, J., & Fothergill, M. (1996). Computer simulation of the triosephosphate isomerase catalyzed reaction. Journal of Biological Chemistry, 271(17), 10010-10016.

[18] Marcus, R. A. (1956). On the theory of oxidation-reduction reactions involving electron transfer. I. The Journal of Chemical Physics, 24(5), 966-978.

[19] Mesecar, A. D., Stoddard, B. L., & Koshland, D. E. (1997). Orbital steering in the catalytic power of enzymes: small structural changes with large catalytic consequences. Science, 277(5323), 202-206.

[20] Pace, C. N., & McGrath, T. (1980). Substrate stabilization of lysozyme to thermal and guanidine hydrochloride denaturation. Journal of Biological Chemistry, 255(9), 3862-3865.

[21] Park, C., & Marqusee, S. (2005). Pulse proteolysis: a simple method for quantitative determination of protein stability and ligand binding. Nature methods, 2(3), 207-212.

[22] Na, Y. R., & Park, C. (2009). Investigating protein unfolding kinetics by pulse proteolysis. Protein Science, 18(2), 268-276.

[23] Kriwacki, R. W., Hengst, L., Tennant, L., Reed, S. I., & Wright, P. E. (1996). Structural studies of p21Waf1/Cip1/Sdi1 in the free and Cdk2-bound state: conformational disorder mediates binding diversity. Proceedings of the National Academy of Sciences, 93(21), 11504-11509.

[24] Bai, Y., Chung, J., Dyson, H. J., & Wright, P. E. (2001). Structural and dynamic characterization of an unfolded state of poplar apo-plastocyanin formed under nondenaturing conditions. Protein Science, 10(5), 1056-1066.

[25] Mullen, G. P., Vaughn, J. B., & Mildvan, A. S. (1993). Sequential proton NMR resonance assignments, circular dichroism, and structural properties of a 50-residue substrate-binding peptide from DNA polymerase I. Archives of biochemistry and biophysics, 301(1), 174-183.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2161v2 | CC BY 4.0 Open Access | rec: 6 Jul 2016, publ: 6 Jul 2016

[26] Uversky, V. N. (2002). Natively unfolded proteins: a point where biology waits for physics. Protein science, 11(4), 739-756.

[27] Vucetic, S., Obradovic, Z., Vacic, V., Radivojac, P., Peng, K., Iakoucheva, L. M., … & Newton, C. D. (2005). Disprot: a database of protein disorder. *Bioinformatics, 21(1), 137-140.*

[28] Wright, P. E., & Dyson, H. J. (1999). Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. Journal of molecular biology, 293(2), 321-331.

[29] Åqvist, J., & Warshel, A. (1993). Simulation of enzyme reactions using valence bond force fields and other hybrid quantum/classical approaches. Chemical reviews, 93(7), 2523-2544.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2161v2 | CC BY 4.0 Open Access | rec: 6 Jul 2016, publ: 6 Jul 2016

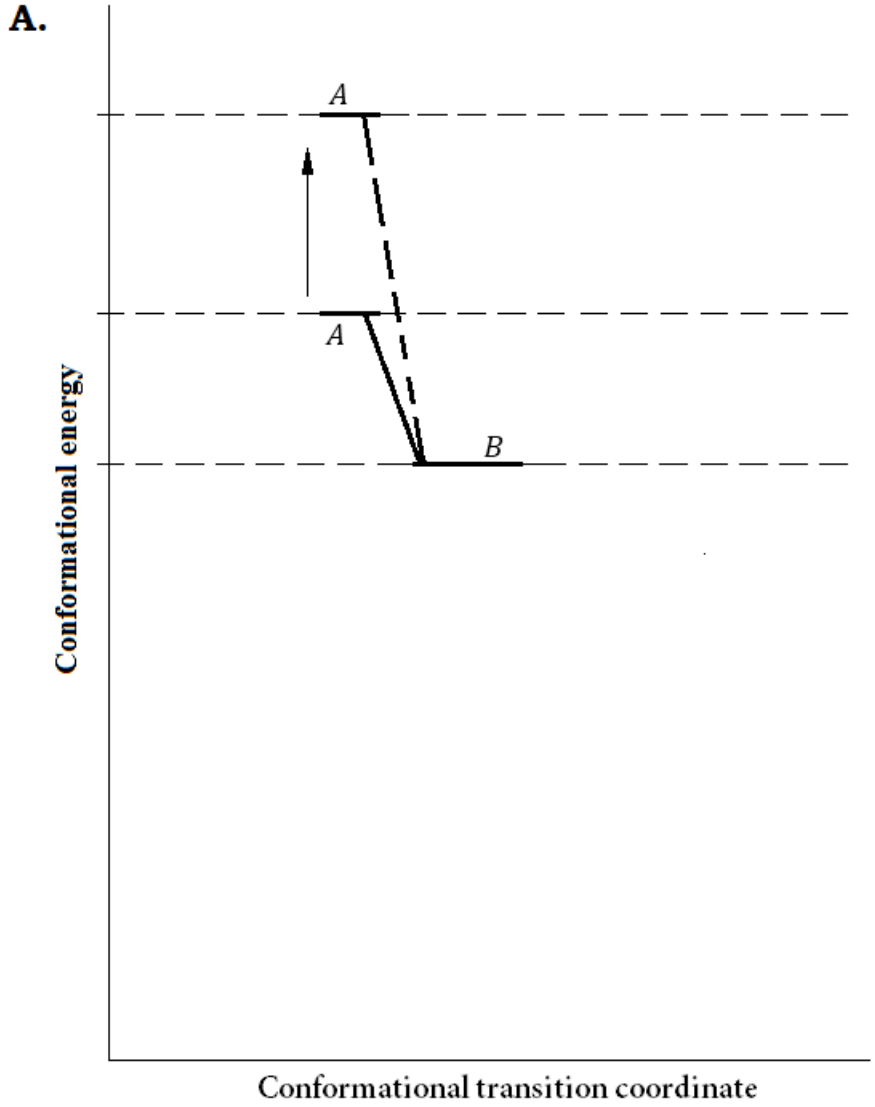

Figure 5A

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2161v2 | CC BY 4.0 Open Access | rec: 6 Jul 2016, publ: 6 Jul 2016

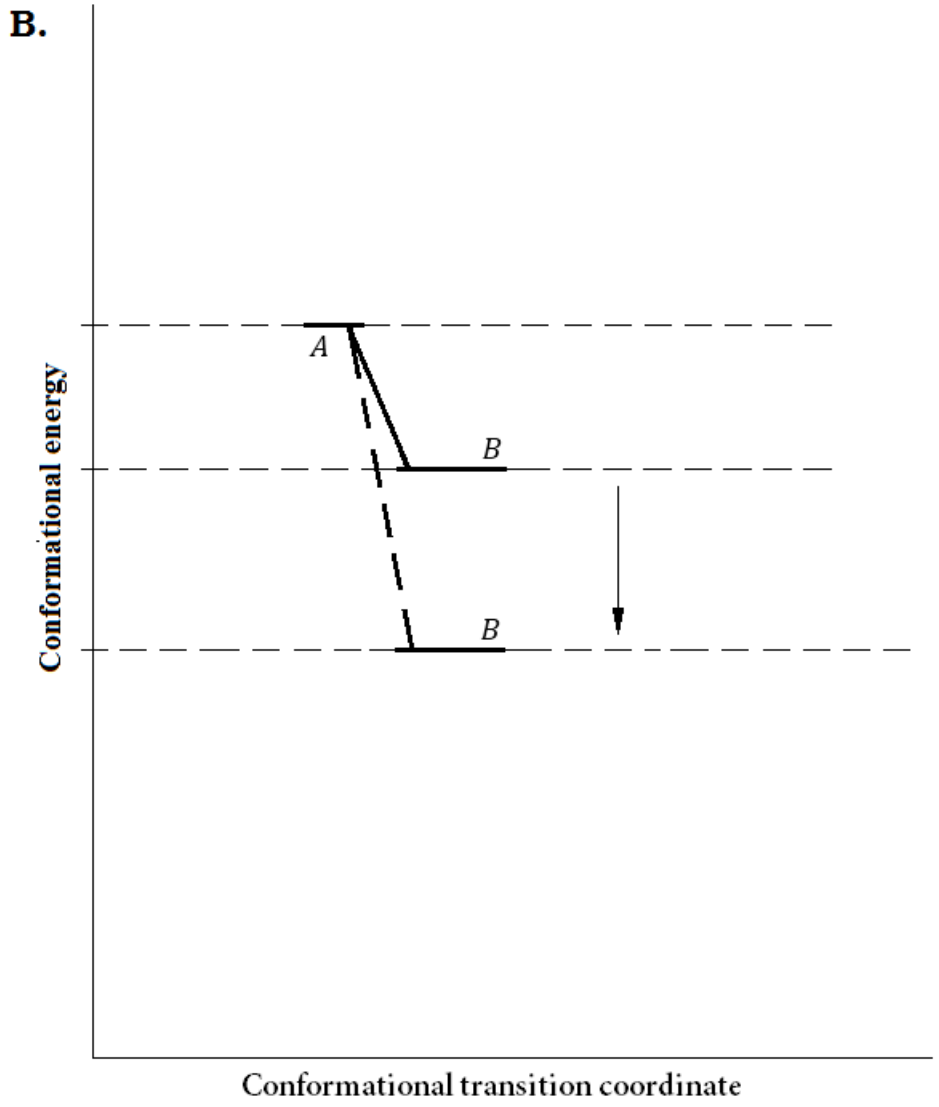

Figure 5B

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2161v2 | CC BY 4.0 Open Access | rec: 6 Jul 2016, publ: 6 Jul 2016

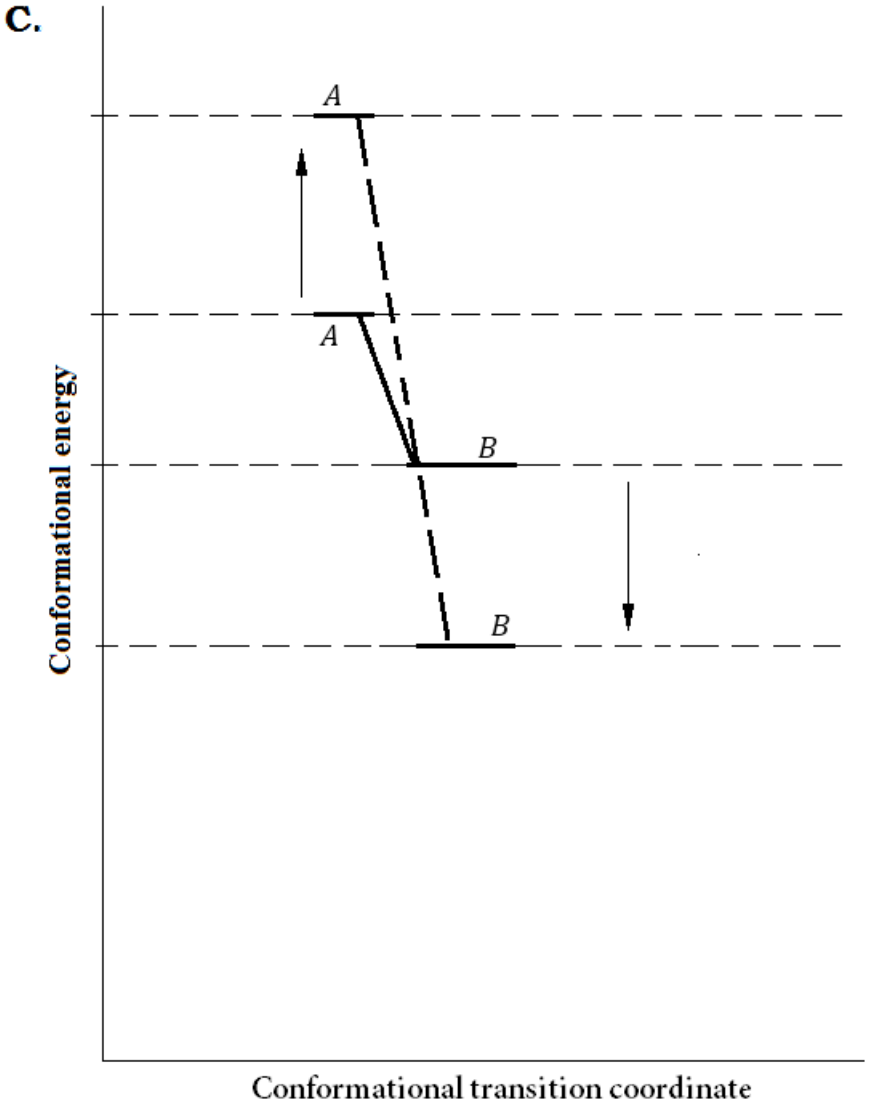

Figure 5C

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2161v2 | CC BY 4.0 Open Access | rec: 6 Jul 2016, publ: 6 Jul 2016

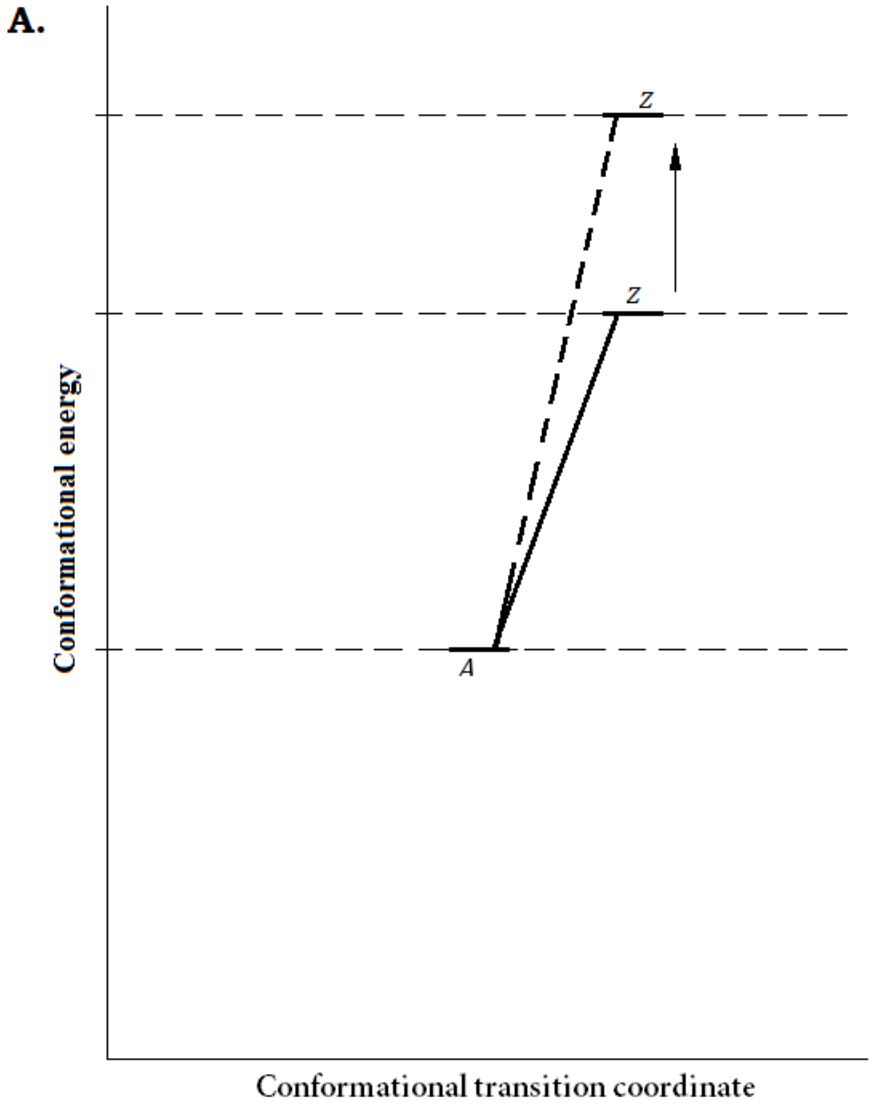

Figure 6A

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2161v2 | CC BY 4.0 Open Access | rec: 6 Jul 2016, publ: 6 Jul 2016

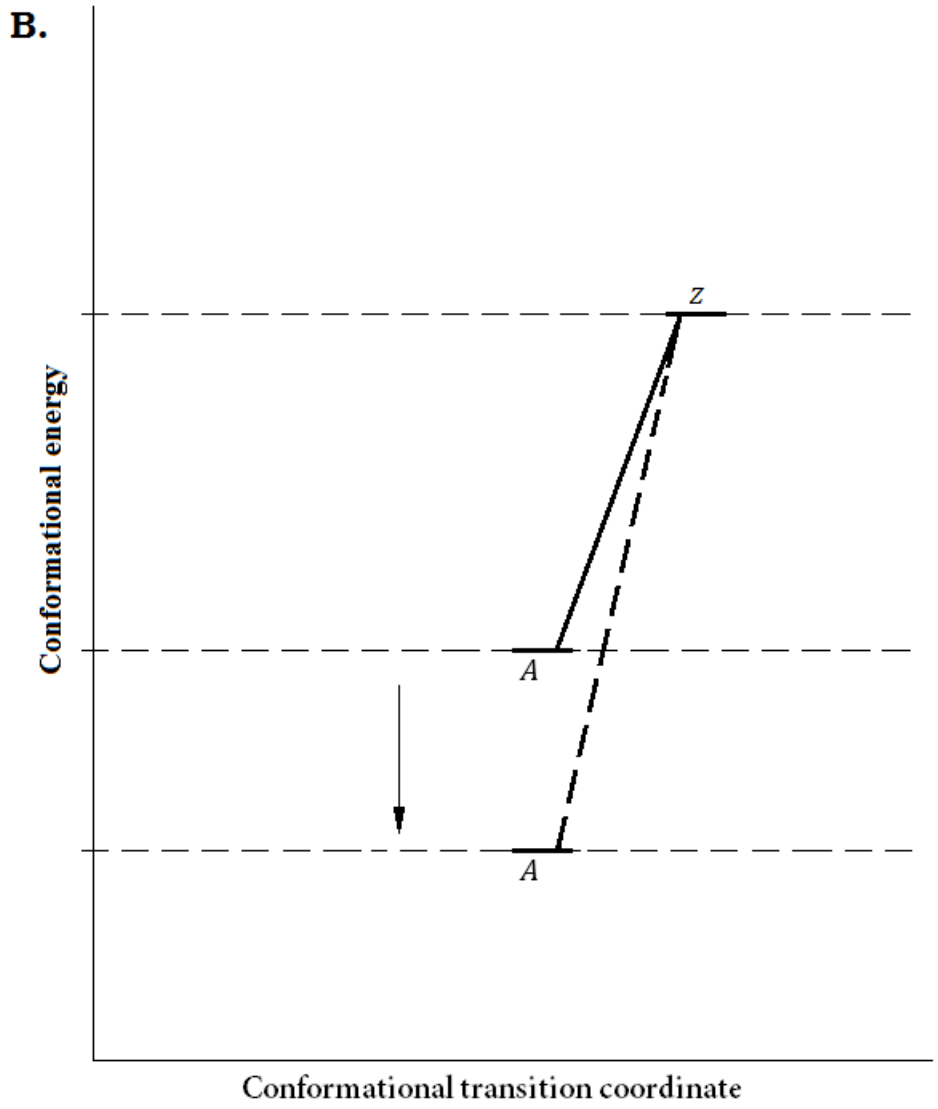

Figure 6B

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2161v2 | CC BY 4.0 Open Access | rec: 6 Jul 2016, publ: 6 Jul 2016

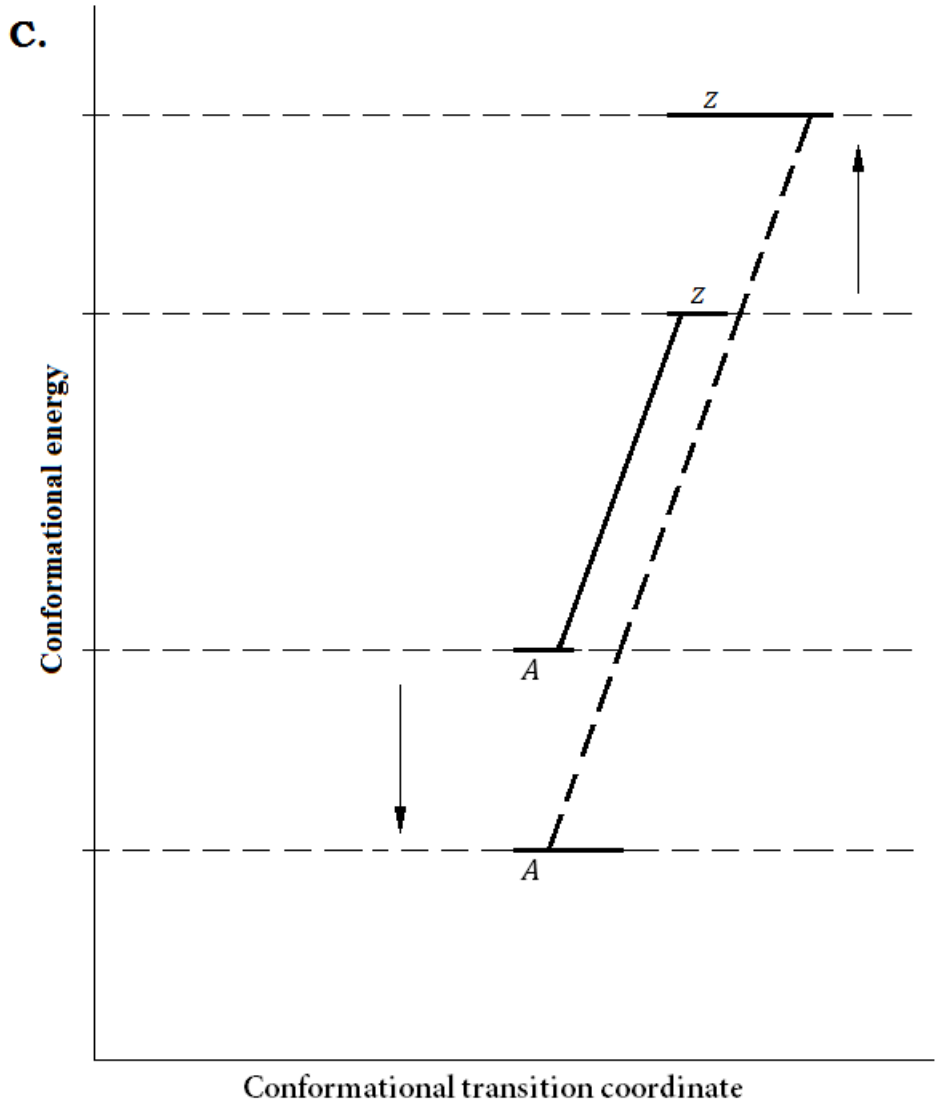

Figure 6C

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2161v2 | CC BY 4.0 Open Access | rec: 6 Jul 2016, publ: 6 Jul 2016