Regular paper

Theoretical model explaining the relationship between the molecular mass and the activation energy of the enzyme revealed by a large-scale analysis of bioinformatics data

Piotr H. Pawlowski[✉] and Piotr Zielenkiewicz

Institute of Biochemistry and Biophysics of the Polish Academy of Sciences, Warszawa, Poland

A general dependence of the enzyme catalytic rate on its mass was revealed when a statistical analysis of 17065 records from the EMP database was performed. The estimated activation energy of the catalytic process decreases asymptotically with the enzyme molecular mass increase. The proposed theoretical model postulates the existence of an intermediate complex of the enzyme and the departing product. It allows for the explanation of the discovered mass-energy relationship, as an effect of the global enzyme–product interactions during complex dissociation. Fitted parameters of the model seem to be in agreement with those widely accepted for the van der Waals energy of molecular interactions. Their values also agree with the picture of the hydrogen bonding in the catalytic process and suggest that surface walk can be the favorable way of the product departure.

INTRODUCTION

Since the Emil Fischer's "lock and key" concept (Fischer, 1984), there were many attempts (Kraut, 1988; Kohen & Klinman, 1998) to explain how the rates of enzyme-catalyzed reactions can be enhanced by a factor of up to 10^{17} compared to the uncatalyzed reaction (Fersht, 1999). Most of them, as the ideas of the Circe effect (Jencks, 1975), orbital steering (Mesecar et al., 1997), and proximity effects (Bolam et al., 1998), incorporated some aspect of entropy. Others show the importance of molecular strain and stretching (Yin et al., 2003) or the role of substrate oscillations (Antoniou & Schwartz, 1997). Original concepts of a low barrier hydrogen bond (Cleland et al., 1998), anti-solvent protection (Cannon & Benkovic, 1998) and electrostatic stabilization (Warshel, 1998) are focused on details of the arrangement of molecules in the active site.

Among those usually specific theories, two general approaches stand out, the ideas of Pauling (Pauling, 1946) and Koshland (Koshland, 1958). The first one suggested that key is not the substrate but a transition state with an activation energy lowered by the enzyme. The second one added that the enzymatic lock undergoes induced changes and only then fits the key. Both the ideas were modified occasionally, disappearing and reappearing in the area of researchers' interest, but they are still fruitful and inspiring.

To test the general "enzymic" ideas, or to look for new ones, one needs large sets of experimental data of a given type. They have finally emerged in the era of bioinformatics owing to the rapid development of databases of metabolic compounds and pathways. Such collections like the now-historical EMP Project (Selkov et al., 1996; Burgard & Maranas, 2001) or BRENDA (Schomburg et al., 2002), KEGG (Ogata et al., 1999), ENZYME (A. Bairoch, 2000), and other (Karp et al., 2000), allow a quick and easy access to the required information. Considering statistical analysis, it is worth stress that EMP database offered a large set of numerical data during a single query session.

Besides its obvious generality, another advantage of statistical analysis is its sensitivity. Statistics can reveal even weak properties or trends among the objects or processes investigated. That was true in our case, when we analyzed over seventeen thousand records from the EMP database, each containing an enzyme's molecular mass and turnover number, and found out that, in general, the activation energy of the catalytic process depends inversely on the molecular mass of the enzyme. Such finding suggests that the discussed energy is influenced by global interactions of the metabolite and the enzyme. In other words, the activation energy includes a contribution from the long-range metabolite-enzyme interactions, which may vary during catalysis. Such claim directly provokes the postulate of a role of relative displacement of molecules during the activation process.

The postulate of displacement differs from the simple concept of a specifically immobilized metabolite, tightly bound in the catalytic site, proposed in the historical "lock and key" model. To overcome the raised discrepancy between the concepts of an immobilized and mobile metabolite, we applied the working hypothesis of an intermediate complex of the enzyme and the departing product. This is a concept borrowed from the model of the chemical kinetic mechanism with central complexes (Leskovac, 2003). According to this hypothesis, metabolites are bound to enzymes not only in the transition state. The products, may initially be bound to the enzymes (Alberty, 1959; Cleland, 1970; Plowman, 1972; Segel, 1993), too. Consequently, we find it reasonable to consider the possible role of an energy barrier to be overcome during the product release (Chaplin & Bucke, 1990; Segel, 1993).

Thus, in the first approximation we postulate that the product release from the catalytic site can take

[✉]e-mail: piotrp@ibb.waw.pl

much more time than the catalysis itself. This is also not a new concept. Already in the classical approach, Michaelis and Menten (Michaelis & Menten, 1913) assumed that the [*SE*] complex breaks down to the free product and the free enzyme much more slowly than it dissociates. Moreover, the product release may be the rate-limiting step (Segel, 1993; Sierks *et al.*, 1997). Then, the turnover number, especially for bimolecular products (Adams & Taylor, 1992) and diffusive processes (Aubol *et al.*, 2004), mainly depends on the rate of the product release, not the rate of the substrate-product transformation. If this is true, the activation energy is not the difference between the Gibbs energies of the transition state and the ground state of the substrate, but rather those of the free and bound product. In this manner, the activation energy can be modified by an essential contribution originated from the change in the energy of global interactions, during the product displacement outside the catalytic site.

Based on the above hypothesis, a theoretical model is presented describing the product release. It takes into account the local and the global enzyme-product interactions. The former ones are treated in a general manner. In the latter case, London forces are considered. One of the conclusions of the model is that the preferred direction of the product release falls tangentially to the enzyme surface. Another prediction, the mathematical formula describing the dependence of the activation energy on the enzyme mass was fitted to the statistical data. Finally, the estimated parameters could be reasonably related to the typical values of: minimal energy of van der Waals interactions, the interatomic distance of the above minimum, specific density of proteins and other. In that way it was found that the magnitude of the tangential displacement of the product along the enzyme surface, during its release, should be of the order of the product size. The analysis of the data dispersion indicates the role of hydrogen bonds, binding the product in the catalytic site.

DATA AND STATISTICAL ANALYSIS

The purpose of this work is to study the rate of catalysis as a function of the enzyme mass. To reach the goal, 17065 records from the Enzymes and Metabolic Pathways (EMP) database were analyzed. The records are presented in supplementary material.

(Supplementary material: http://www.ibb.waw.pl/~piotrp/SUPP_MAT/data.xlsx).

The search field included the following areas: Enzyme Commission Number (EC), Reaction (*RE*), Molecular Mass (M_{mE}), and Catalytic Constant (k_c), where: k_c denotes the turnover number of the catalytic center in [$1/s$], as in Table 1.
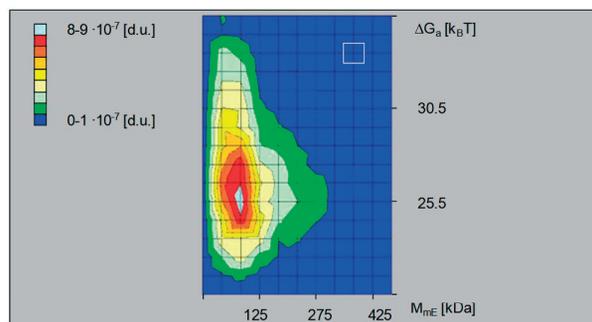
For each different record, the value of the Gibbs energy of activation, ΔG_a, was attributed according to the Eyring formula (Eyring, 1935):

$$\Delta G_a = -\,k_B T \ln\!\left(\frac{k_c h}{k_B T}\right) \tag{1}$$

where: k_B is the Boltzmann constant, T is the absolute temperature and h is the Planck constant. In the following, it is assumed $T = 300$ [K].

Next, each group of records with the same EC, *RE*, and M_{mE}, but usually dispersed values of ΔG_a, were replaced by a single statistical representative. Each representative was described by the pair of values M_{mE} and the average value of ΔG_a. Thus $N = 6915$ different, original or averaged, mass-energy records were obtained and further analyzed.

A contour chart of the probability density of an enzyme's occurrence in the two-dimensional space $M_{mE} \times \Delta G_a$ (Fig. 1) was constructed. For this purpose, representative records were counted in probing areas, 1 [$k_B T$] × 50 [kDa], and the obtained numbers were divided by $5 \cdot 10^4$. The resulting densities, labeled by the coordinates of the center of the probing area, were visualized on an Excel surface chart.

Table 1. An example of records from the EMP database*

EC	RE	M_{mE}	k_c
1.1.0.0	17'b-estradiol + NAD('+) = estrone + NADH[**]	107692	5.67
1.1.0.0	17'b-estradiol + NAD('+) = estrone + NADH[**]	107692	0.631
1.1.0.0	17'b-estradiol + NAD('+) = estrone + NADH[**]	107692	0.587
1.1.0.0	17'b-estradiol + NAD('+) = estrone + NADH[**]	108000	5.67
1.1.0.0	17'b-estradiol + NAD('+) = estrone + NADH[**]	108000	0.631
1.1.0.0	17'b-estradiol + NAD('+) = estrone + NADH[**]	108000	0.587
1.1.1.0	(`R`)-lactaldehyde + NAD(P)('+) = methylglyoxal + NAD(P)H	42000	97.2
1.1.1.0	(`R`)-lactaldehyde + NAD(P)('+) = methylglyoxal + NAD(P)H	43000	97.2
1.1.1.0	(`R`)-mandelate + NAD('+) = benzoylformate + NADH [**]	77200	4750
1.1.1.0	`sn`-glycerol 1-phosphate + NAD(P)('+) = dihydroxyacetonephosphate + NAD(P)H	302000	25800
1.1.1.0	2,5-diketo-D-gluconate + NADPH = 5-keto-D-gluconate + NADP('+) [**]	70000	1020
1.1.1.0	2,5-diketo-D-gluconate + NADPH = 5-keto-D-gluconate + NADP('+) [**]	72000	1020
1.1.1.0	2,5-diketo-D-gluconic acid + NADpH = 2-keto-L-gulonic acid + NADP('+) [**]	29000	5070
1.1.1.0	2,5-diketo-D-gluconic acid + NADpH = 2-keto-L-gulonic acid + NADP('+) [**]	34000	504

*column captions:, *EC*, Enzyme Commission Number, *RE*, Reaction, M_{mE}, Molecular Mass [Da], and k_c, Catalytic Constant [1/s]],.

Figure 1. Density contour chart of the probability of enzyme occurrence.
Representative 6915 records (see text) were analyzed. The density unit is [d.u.] = [$(k_B T \, \text{Da})^{-1}$]. The white square indicates dimensions of the probing area.

In the described chart, one can observe a weak symmetry of the results relative to the axis near $\Delta G_a = 25.5$ [$k_B T$] when the molecular mass is greater than 125 [kDa]. For smaller enzymes, this symmetry is broken. Here, an increase in probability of the higher energies was found. This may suggest a relationship between mass and energy.

Furthermore, marginal histograms of ΔG_a (mass neglected) and M_{mE} (energy neglected) were created (Fig. 2a, b). Consistently, the class intervals were 1 [$k_B T$] and 50 [kDa].

As can be seen in Fig. 2a, the marginal distribution of activation energy is bell-shaped. This shape is slightly distorted compared with the fitted normal distribution. It exhibits a small decrease (increase) in the occurrence of energy below (above) the average. This is the effect of the broken symmetry observed in the contour chart (Fig. 1). The estimated average value of the fitted normal distribution, $\Delta \overline{G}_a = 26.59 \pm 0.10$ [$k_B T$], falls in the range of the most common energies, 26–27 [$k_B T$]. Standard deviation $\sigma_G = 3.15 \pm 0.08$ [$k_B T$].
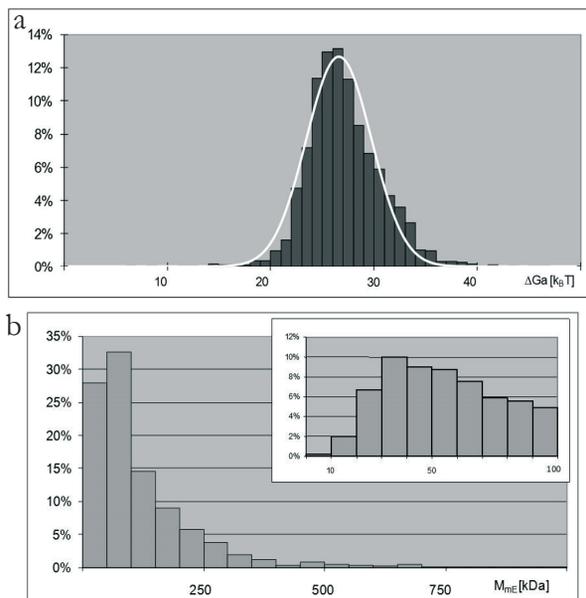
Figure 2. Marginal histograms.
(**a**) Marginal distribution of Gibbs energy of activation ΔG_a. The mean value $\Delta \overline{G}_a = 26.59 \pm 0.10$ [$k_B T$] and the standard deviation $\sigma_G = 3.15 \pm 0.08$ [$k_B T$]. The continuous line shows normal distribution. (**b**) Marginal distribution of the enzyme molecular mass M_{mE}. The mean value $\overline{M}_{mE} = 125.9 \pm 2.3$ [kDa] and the standard deviation $\sigma_M = 187.1$ [kDa]. The inset contains the distribution profile for small molecular masses at the class interval of 10 [kDa].
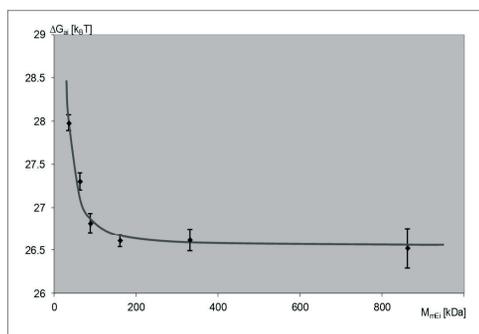
Figure 3. Group mean value of Gibbs energy of activation $\Delta \overline{G}_i$ versus calculated group mean value of mass $\overline{M}_{mEi}$.
Vertical bars show standard errors of estimation. The continuous line indicates numerical fitting of the proposed theoretical function: $\Delta \overline{G}_i (\overline{M}_{mEi}) = a + \dfrac{b}{\overline{M}_{mEi}^{5/3}}$, where are the determined parameters (see Results).

The marginal distribution of molecular mass, Fig. 2b, seems to be more skewed than the marginal distribution of energy. This shape is similar to the shape of a cut in the contour chart (Fig. 1), perpendicular to the vertical axis near the point $\overline{M}_{mE} = 75$ [kDa], $\Delta G_a = 25.5$ [$k_B T$]. The estimated mean $\overline{M}_{mE} = 125 \pm 2.3$ [kDa] falls outside the range of the most common values, 50–100 [kDa]. Standard deviation $\sigma_G = 187.1$ [kDa].

To answer the question whether the Gibbs energy of activation significantly depends on the molecular mass of the enzyme a single-factor analysis of variance was performed. Representative records were divided into six classes of mass magnitude: $M_{mE} \leq 50$ [kDa], $50 < M_{mE} \leq 75$ [kDa], $75 < M_{mE} \leq 100$ [kDa], $100 < M_{mE} \leq 250$ [kDa], $250 < M_{mE} \leq 500$ [kDa], and $M_{mE} > 500$ [kDa]. The number of analyzed enzymes in those classes was, respectively: 1928, 1306, 954, 2021, 553, and 153. The separated groups 1–4 derived from the "main body" of the distribution of mass, and 5–6 from the distribution tail. Then, mean group values of energy, $\Delta \overline{G}_{ai}$, and mean group values of mass, $\overline{M}_{mEi}$, ($i = 1, 2, ...6$) were calculated. Next, the F-statistic value was calculated concerning respective energy variances. The obtained value $F = 36$, at 5 and 6909 degrees of freedom, allows us to reject the hypothesis that $\Delta \overline{G}_{ai}$ are equal to each other. The probability of the null hypothesis is less than $3 \cdot 10^{-36}$.

To analyze the relationship between ΔG_a and M_{mE}, the mean group values $\Delta \overline{G}_{ai}$ were plotted versus the mean group values $\overline{M}_{mEi}$. The result in Fig. 3 clearly shows an asymptotic decrease of the activation energy with the increase in the molecular mass of the enzyme. A quantitative explanation of this phenomenon is proposed below. It is based on a developed theoretical model describing the interactions of enzyme and product.

THEORETICAL MODEL

Let us consider the simple enzymatic process

$$S \xrightarrow{E} P \tag{2}$$

consisting of three elementary steps, i.e., the binding of the substrate to the active site of the enzyme,

$$S + E \rightarrow [SE] \tag{3}$$

the transition of the substrate to the product,

$$[SE] \rightarrow [PE] \tag{4}$$

and the release of the product,

$$[PE] \rightarrow P + E \tag{5}.$$

For such a process, the total time taken by a single turnover of the enzyme (eqn. 4 and eqn. 5) can be generally estimated as

$$\tau = \tau_1 + \tau_2 \tag{6}$$

where τ_1 is the time taken for the substrate-product transition in the active site (eqn. 4), and τ_2 is the time taken for the product release (eqn. 5).

The turnover number k_c, thanks to the definition

$$k_c = \frac{1}{\tau} \tag{7}$$

(single catalytic site was assumed) can be expressed as

$$k_c = \frac{1}{\tau_1 + \tau_2} \tag{8}.$$

The model of the "ideal enzyme" considered here allows for the acceptance of the working assumption that pure catalysis (substrate-product transformation) is a faster step than the product departure, i.e., time τ_1 is much shorter than τ_2. Thus, approximately

$$k_c = \frac{1}{\tau_2} \tag{9}.$$

Let us look at the product release from the Arrhenius perspective, i.e., as a kind of an attempt to overcome a certain hypothetical energy barrier, namely ΔG_a. This attempt is repeated with the frequency ν and the probability of success p, where:

$$\nu = \frac{k_B T}{h} \tag{10}$$

and

$$p = exp(-\frac{\Delta G_a}{k_B T}) \tag{11}.$$

One can also state that

$$\frac{1}{\tau_2} = \nu p \tag{12}.$$

Finally, the set of eqs. 9–12 leads to the relationship

$$k_c = \frac{k_B T}{h} exp(-\frac{\Delta G_a}{k_B T}) \tag{13}.$$

The above Eyring type formula relates the turnover number k_c with the activation energy ΔG_a of the product release. From a thermodynamic point of view, the activation energy ΔG_a is the difference in useful energies of the system, strictly, the energy of the state preceding the product release and the energy of the intermediate complex $[PE]$. It is convenient to label the characteristic states of the system by appropriate indexes "b" (bound) and "r" (release), as in Fig. 4.

Estimation of the discussed energy requires additional thermodynamic analysis of the interactions between enzyme and product. To do that, let us consider an enzyme and a product as a small system consisting of the
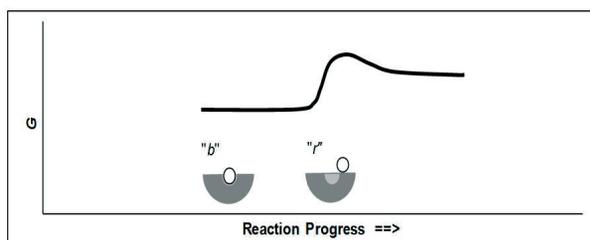
Figure 4. Change in useful energy, G, of the enzyme-product-interaction field system along with reaction progress.
Bottom, cartoons "b" and "r", show the system respectively in the bound-state of the enzyme–product complex and when the system crosses the barrier for product release.

two subsystems and a conservative field of the enzyme-product interaction, surrounded by a thermostatic environment. The useful energy G of such a system can be expressed as

$$G = G_E + G_P + U_{Field} \tag{14}$$

where G_E and G_P are the Gibbs free energies of enzyme and product, respectively, and U_{Field} is the potential energy of the enzyme-product interaction field.

The activation energy ΔG_a of product release can be estimated as the difference between the useful energies of the system in states r and b. Thus,

$$\Delta G_a = G_E^r + G_P^r + U_{Field}^r - (G_P^b + G_P^b U_{Field}^b) \tag{15}.$$

At a constant pressure and temperature, we can assume that neither enzyme nor product does not change during the process of product release. Thus,

$$G_E^b = G_E^r, \tag{16}$$

$$G_P^b = G_P^r. \tag{17}$$

Taking into account the above assumption (eqns. 16, 17), the activation energy (eqn. 15) can be estimated as

$$\Delta G_a = U_{Field}^r - U_{Field}^b. \tag{18}.$$

To describe the energy of the interaction field, U_{Field}, assume that the enzyme-product interactions can be divided into two different groups, local interactions (loc), and global interactions (glo). By their nature, the local and the global interactions are respectively driven by short-range, and long-range forces. Due to the above,

$$U_{Field} = U_{loc} + U_{glo} \tag{19}.$$

Now one can express the activation energy (eqn. 18) as a superposition of the two components manifesting the local and the global interactions of the enzyme and the product. Then we get, respectively,

$$\Delta G_a = \Delta U_{loc} + \Delta U_{glo} \tag{20}$$

where

$$\Delta U_{loc} = U_{loc}^r - U_{loc}^b \tag{21}$$

$$\Delta U_{glo} = U_{glo}^r - U_{glo}^b \tag{22}.$$

To investigate the activation energy dependence on the enzyme size let us focus our attention on the global component ΔU_{glo}. Microscopically, the discussed input is the effect of the long-range interactions of the atoms of the enzyme and the atoms of the product. The energy u_{ij} of such interactions has been proposed to be described by the London forces (London, 1937). So,

$$u_{ij} = -2e_{ij}\left(\frac{d_{ij}^*}{d_{ij}}\right)^6 \tag{23}$$

where, d_{ij} is the distance between the i-th atom of the enzyme and the j-th atom of the product, e_{ij} is an absolute value of the minimal energy of van der Waals interactions, and d_{ij}^* is the interatomic distance at the energy minimum.

As the energy of the enzyme-product global interactions, U_{glo}, is the sum of all binary inputs u_{ij}, i.e.,

$$U_{glo} = \sum_{ij} u_{ij} \tag{24}$$

the input of the global interactions in the activation energy (eqn. 22) can be calculated as:

$$\Delta U_{glo} = \sum_{ij} u_{ij}^r - \sum_{ij} u_{ij}^b \qquad (25).$$

Specific computation of the above equation requires a detailed description of the architecture of the system and the localization of the process.

For simplicity, let us consider interacting product and enzyme as a relatively small spherical particle and a large hemisphere. Let us also assume that the bound-state product initially sits (half-immersed) in the center of the hemisphere (the center of the great circle), and while the state of the system changes to the state before the product release, the product displaces over the enzyme surface to a certain new location near the center. The assumed geometry is shown in Fig. 5, where R_E is the radius of the enzyme, and R_P is the radius of the product.

To simplify the whole model, homogeneity of the enzyme and the product is assumed, represented by constant densities of the number of atoms n_E and n_P, respectively. To set the required parameters, the London interactions between all the atoms of enzyme and product are assumed to be the same, which allows us to replace the parameters e_{ij} and d_{ij}^* by a single pair of the independent constants: e and d^*.

At the above assumptions, mathematical analysis presented in Appendix 1 (http://www.actabp.pl/pdf/2_2013/239-suppl.pdf) shows that

$$\Delta U_{glo} = \frac{16}{3} \pi^2 e (d^*)^6 n_E n_P \frac{R_P^3 \zeta^2}{R_E^5} \qquad (26)$$

where ζ is the distance between bound-state site and the site of the hypothetical barrier for the product release.

Finally, eqn. 20 and 26 lead to the equation

$$\Delta G_a = \Delta U_{loc} + \frac{16}{3} \pi^2 e (d^*)^6 n_E n_P \frac{R_P^3 \zeta^2}{R_E^5} \qquad (27).$$

RESULTS

Predictions of the proposed *theoretical model* (eqn. 27) allow modeling of the dependence of activation energy ΔG_a on enzyme mass. To reach the goal the molecular
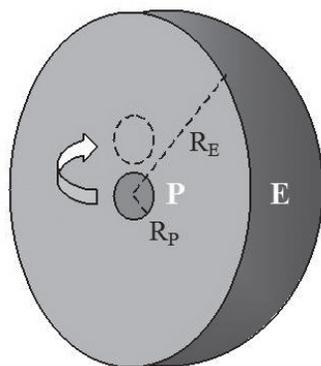
Figure 5. Geometrical assumptions of the model.
A small spherical particle of the bound product (P) in the bound-state is placed in the center of the big hemispherical enzyme (E). Geometrical localization of the point when the transition of the system to the state of the product release takes place is in close neighborhood of the center (broken-line circle). Symbols denote: R_E, radius of enzyme, R_P, radius of product.

masses of the enzyme and the product, M_{mE}, M_{ml} were approximated by equations:

$$M_{mE} = \frac{2}{3} \pi R_E^3 n_E \mu_E \qquad (28)$$

and

$$M_{mE} = \frac{4}{3} \pi R_P^3 n_P \mu_P \qquad (29)$$

where μ_E and μ_P are the mean atomic masses of the atoms of enzyme and product, respectively.

For convenience, the densities of the number of atoms were expressed as:

$$n_E = \frac{\varrho_E}{\mu_E m_U} \qquad (30)$$

$$n_E = \frac{\varrho_P}{\mu_P m_U} \qquad (31)$$

where ϱ_E and ϱ_P are specific masses of the enzyme and the product. m_U is the atomic mass unit.

For clarity, we propose to describe the pre-release displacement ζ using the product radius as a length unit. So,

$$\zeta = \alpha R_P \qquad (32)$$

where α is a dimensionless parameter.

Then, the analyzed energy (eqn. 27) can be expressed as a function of the enzyme mass,

$$\Delta G_a = \Delta U_{loc} + \frac{\Phi}{M_{mE}^{5/3}} \qquad (33)$$

where coefficient Φ is defined as

$$\Phi = \frac{16 \pi^2 \alpha^2 e (d^*)^6 \rho_E^{8/3} M_{mP}^{5/3}}{3 \mu_E \mu_P m_U^2 \rho_P^{2/3}} \qquad (34).$$

Assuming eqn. 33 as a model of nonlinear regression and then fitting the hypothetical function $\Delta \overline{G}_i (\overline{M}_{mEi})$ = a + b/$\overline{M}_{mEi}^{5/3}$ to the statistical data (Fig. 3), we obtain $a = \Delta U_{loc} = 26.55 \pm 0.06$ [$k_B T$], and $b = \Phi = (5.52 \pm 0.47) \cdot 10^7$ [$k_B T$ Da$^{5/3}$]. Additionally, assuming that $e = 0.25$ [$k_B T$] and d$^* = 4 \cdot 10^{-10}$ (m) (as for C-C atoms in AUTODOCK database (Morris *et al.*, 1998)), $\varrho_E = \varrho_P = 1.33 \cdot 10^3$ [kg/m^3] (San Martín, *et al.*, 1995), $M_{mP} = 350$ [Da], $\mu_E = \mu_P = 10$ [Da] and $m_U = 1.66 \cdot 10^{-27}$, we estimate $\alpha = 3.03$.

Similarly, the energy of global interactions in the enzyme center, $U_{glo}(0)$, (see A.1.8) can be expressed as

$$U_{glo}(0) = \frac{16 \pi^2 e (d^*)^6 \rho_E \rho_P}{9 \mu_E \mu_P m_U^2} \left[\frac{1}{2} \frac{M_{mP} \rho_E}{M_{mE} \rho_P} - 1 \right] \qquad (35).$$

Taking $M_{mP} = 50\,000$ [Da] and other parameters as above, we obtain $U_{glo}(0) = -115$ [$k_B T$].

DISCUSSION

In a discussion of enzymatic catalysis phenomena mechanisms of pure catalysis usually take central stage. It is obvious when investigatig catalysis as such, but it is not so obvious when discussing its kinetics, and the process of product release (Segel, 1993). The accompanying non-catalytic conformational changes and the overall relative movement of the interacting partners during substrate binding or product release can influence the effective rate of the total process, so they should also be seriously considered (Lerch *et al.*, 2002; Tian *et al.*, 2002).

In our opinion, the large set of presented statistical data of activation energy, calculated from the turnover number, give us an argument to think about the effective role of such a nonspecific processes in catalysis. The observed tendency of a systematic increase in the probability of occurrence of higher activation energies for enzymes of smaller molecular mass (Fig. 1), supported by the analysis of variance and the mean group values picture (Fig. 3), clearly shows that there is a systematic mass-energy dependence. According to the averaged picture, the activation energy asymptotically decreases with an increase in the molecular mass of the enzyme. The documented decrease reaches approximately, 1.5 $[k_BT]$ i.e., 6% of the final level.

The activation energy by definition answers to the change in energy of intermolecular interactions (Willemsen et al., 2000). The value $[k_BT]$ of the decrease is smaller than a typical energy 1.7–3.7 $[k_BT]$ of a hydrogen bond (Myers & Pace, 1996), and higher than the minimum energy of van der Waals forces 0.17–0.34 $[k_BT]$ (Levitt & Perutz 1988). This suggests that many interactions of the second type can be involved in the discussed phenomenon.

Further, it is difficult to expect that the change in the energy of local interactions in the catalytic site could depend on the entire enzyme mass. Thus, the explanation of the observed tendency offered by our hypothesis seems to be reasonable. It states that during the enzymatic process, there is not only a local, enzyme-metabolite interplay, but also a global one involving the entire components of complex. As the main known candidates for the global players, i.e., long-range interactions, are dependent on the relative position, their input into activation energy requires relative displacement of the reactants. The best opportunity for such a displacement takes place when the bound product is released. This is what our theoretical model postulates. It relates the turnover number with the frequency of successful product departure from the catalytic site, and the activation energy with the energy barrier to be overcome during the product release.

In the first approximation our model neglects the relatively short duration of the step of pure catalysis (Lerch et al., 2002). By that means it omits the contribution to the catalytic constant from the ealier steps of the catalytic process (Leskovac, 2003). The ideas of the bound product and of the last step lowering the overall rate have been previously described in the literature (Michaelis & Menten, 1913; Alberty, 1959; Cleland, 1970; Plowman, 1972; Adams & Taylor, 1992; Segel, 1993; Sierks et al., 1997; Aubol et al., 2004), especially in the kinetic models considering several central complexes and an ordered sequential product release. Although there are very well known reports of product dissociation as the rate- limiting step (Narayana et al., 1997; Stoddard et al., 1998), the picture of the enzyme as a slow "releaser" is not commonly accepted.

Despite the above, concepts mentioned are the foundation of our "ideal enzyme" approach, which substantially basically differs from the popular one, but is an effective tool in looking for a manifestation of noncatalytic effects in the vast sets of experimental data. In the popular approach, the activation energy is treated as a jump in the energy profile along the pathway of the substrate to product conversion (Huml, 1999), and thus the final process of the product release is neglected. In our proposal, the main energy barrier determining the rate of the process is not formed by the action of chemical bonds in the substrate. Instead, it is created by local interactions of the enzyme and product, modified by the global London forces.

An important argument for proposed concept and related model is the successful fit of the predicted mass-energy function to the set of averaged experimental data (Fig. 3).

Furthermore, the physical parameters of the model can be reasonably adjusted (see Results) to the typical minimum energy of van der Waals interactions, the minimum energy distance, the density of protein, the molecular mass of the product, and the mean molecular mass of the atom, giving a prediction of the surface displacement of the product $\varrho = 3.03 \, R_P$, where R_P is the radius of the product.

A displacement of a magnitude of approximately three product radii should probably fully uncover the catalytic site for a new substrate. This defines quantitatively the meaning of the term "product release", and where the geometrical trajectory of the full catalytic process ends.

This also shows that the concept of product release is not identical with the concept of the free product far from the catalytic site (or the enzyme surface). It rather means such a position of the product on the enzyme surface which fully exposes a free catalytic site for a new substrate and does not influence the rate of their interaction.

Let us now discuss the main assumptions of the model, i.e., the idealization of the enzyme function, its hemispherical shape and the surface displacement of the released product.

The working assumption of a super-fast step of the catalytic process makes activation energy related to pure catalysis apparently neglected, which can lead to some kind of misinterpretation. In fact the part of the energy of the local interactions could be attributed to the omitted energy of the pure process, and is planning to include it in a future more detailed model.

We chose a hemisphere because it is the simplest geometric form which has a focusing property, which means that the long-range London forces attract every molecular body to some points on its surface. This is the center of the great circle and the opposite point on the axis of symmetry. They are both good candidates for localization of the model catalytic site. We chose the first one as energetically favorable (on average, the enzyme body is closer to the bound product).

Furthermore, the surface displacement of the product is energetically favorable compared to the displacement in the perpendicular direction. In the hemisphere center, the tangential component of acting force vanishes, but not the perpendicular one. The concept of the substrate shift on the enzyme surface (Tian et al., 2002), from a docking site to a catalytic one, can be treated as a precursor of our idea. Furthermore, special routes for the product regress were indicated in the work of Schleinkofer (Schleinkofer et al., 2005).

On the other hand, the typical total energy of product binding can be estimated as $\Delta U_{loc} + |U_{glo}(0)| \approx 140 \, [k_BT]$ (see Results). Now, it is obvious that it is impossible for a bound molecule to shift from the catalytic site to "infinity" during one thermal jump. As we believe, the process of product moving away requires some steps, or "climbing along the mountain path of energy", on the enzyme surface. Probably the estimated activation energy ΔG_a is only an increase in the energy during the first step on this pathway. Unpublished calculations show that the energy needed to overcome global interactions and launch the product to infinity is approximately four times smaller for a product located on the enzyme great

circle than for a product in the enzyme center. Then, the "escape" energy may be comparable to that needed during the process of leaving the catalytic site.

It should be added that the main feature of other assumptions is their simplicity. These are: single-product kinetics, homogeneity of interacting molecules, the spherical shape of the product, and constant mass. This is also a simplification neglecting the multi-domain protein structures and the special role of globular parts. They have to be treated as a compromise between the diversity of the data set and the generality of mathematical considerations.

Another argument in favor of our hypothesis comes from the analysis of the main characteristics of the activation energy distribution (Fig. 2a). If one assumes that the number of bonds of a certain type that have to be broken during the catalytic process is governed by Poisson statistics, then the expected standard deviation of energy, σ_G, should be equal to:

$$\sigma_G = e_b \, \mathrm{Sqrt}(n_b) \tag{36}$$

where n_b is the mean number of bonds, and e_b is the bond energy (see Appendix 2 at http://www.actabp.pl/pdf/2_2013/239-suppl.pdf).

Now, if one assumes that the discussed energy is the change in Gibbs energy only during the pure catalysis, should be of the order of the energy of a single covalent bond ($n_b \approx 1$). This is not so in our case, where the standard deviation $\sigma_G = 3.15 \, [k_B T]$ is much smaller than the energy of the order of 100 $[k_B T]$, typical for a single covalent bond (Pauling, 1960).

The conclusion will be the same if we assume that in the transition state the bond energy, e_b, is lowered by the enzyme to the level of the average activation energy (26.59 $[k_B T]$). This supports our hypothesis according to which we are not dealing with pure catalysis.

In the following discussion we try to answer the arising question of what type of local forces determine the discussed energy. To do that, let us look closely at the average energy. The mean value $\Delta \overline{G}_a = 26.59 \, [k_B T]$ differs slightly from the estimated value of the increase in the energy of local interactions, $\Delta \overline{U}_{loc} = 26.55 \, [k_B T]$. This difference is mainly due to the increase in energy of global interactions, appearing in low-molecular-mass enzymes. Neglecting the mass effect on the distribution characteristics, we can assume that the observed distribution describes the probability of an increase in energy of local interactions, and then we can estimate (see Appendix 2 at http://www.actabp.pl/pdf/2_2013/239-suppl.pdf) the energy of a single bond to be overcome. It is

$$e_b = \frac{\sigma_G^2}{\Delta U_{loc}} \tag{37}$$

and the number of bonds is

$$n_b = \frac{\Delta U_{loc}}{e_b} \tag{38}.$$

The proposed estimation gives the values $e_b = 0.37 \, [k_B T]$ and $n_b = 81$. The obtained value of energy is a little higher than the energy of typical van der Waals interatomic forces, and smaller than the energy of a hydrogen bond. One possible explanation can be that local binding is mainly the result of surface van der Waals interactions (Keesom, Debye, London) and some hydrogen bonds. Thus, simply represents an effectively shifted value. To check this possibility let us subtract some contributions from hydrogen bonds: $n_{Hb}e_{Hb}$, $n_{Hb}e_{Hb}$, from the observed dispersion and the mean value. The recalculation

$$e_b = \frac{\sigma_G^2 - n_{Hb}e_{Hb}^2}{\Delta U_{loc} - n_{Hb}e_{Hb}} \tag{39}$$

and

$$n_b = \frac{\Delta U_{loc} - n_{Hb}e_{Hb}}{e_b} \tag{40}$$

where we assume n_{Hb} and $e_{Hb} = 1 [k_B T]$ gives us a reasonable value $n_b = 101$, for van der Waals interactions.

The above example shows that only ca. 15% of the increase in the energy of local interactions can be due to the breakage of stretched hydrogen bonding, and the dominant rest is due to the breaking of surface van der Waals interactions. It is widely accepted that hydrogen bonds can bind the substrate to the catalytic site. In the light of the above results, their collective interaction with surface van der Waals forces can also bind the product before its release. Such conclusion is an important addition to our hypothesis.

A more general remark can be made following a simultaneous inspection of the distribution of enzyme molecular masses (Fig. 2b) and the regression line in the mean group value plot (Fig. 3). It is related to the frequently asked question of why the enzymes are so big (Mattey et al., 1998).

According to Britt (1997), "for enzymes, bigger is better". He argues that an increased enzyme/substrate mass ratio implies a greater stabilization of the transition state and an increased substrate binding energy. His Shifting Specificity Model (Britt, 1993), relates the strong interaction of the substrate with the enzyme to the modulation of the specific active site transformation and, in this way, to a more rapid conversion of substrate to product(s).

Although we cannot answer exactly the same question, we can take into account the product-enzyme interactions and try to speculate why the enzymes are not small. As it is seen in the inset in Fig. 2, the smallest enzyme mass reported is close to $MM = 5$ [kDa]. If one uses the above mass in the calculation of the activation energy according to the regression model (eqn. 33), for $\Delta U_{loc} = 26.55 \, [k_B T]$ and $\Phi = 5.52 \cdot 10^7 \, [k_B T \, Da]$, one obtains $\Delta G_a \approx 64 \, [k_B T]$. Such a high value of the threshold of energy that needs to be overcome makes the rate-limiting step of product release too much "limiting". Respectively, the average time of product release reaches an impossible value of $\tau_2 = 10^{15} \, [s]$. Simply saying, such small enzymes are expected to be permanently "glued" to the product and thus inoperational, why then some of them do work despite the above? It is plausible that they are characterized by the energy ΔU_{loc} smaller than the global mean for population, assumed in the above estimation.

The above example clearly documents the biological importance of the mass-energy dependence phenomenon. As it was shown, smaller enzymes generally bind the product more strongly than do larger ones. Other consequences of that relation are still under study. For example, it is clearly seen that the mass-energy dependence effect starts to dominate in the range of the most common molecular masses of enzymes (Fig. 2b and Fig. 3). Does it signal some biological consequences of that phenomenon? If so, what may be the profit from some lowering of the enzyme rate? Is it an increased infallibility of the system, decreased crowd or lowered some costs? If not, why was it tolerated during evolution and did not shift to the maximum of mass distribution? Maybe it is still changing. Histogram dispersion is relatively wide.

Considering possible predictions of our hypothesis, we can expect that bigger products are more strongly bond to the enzyme, so a higher activation energy should correlate with a higher molecular mass of the product. To
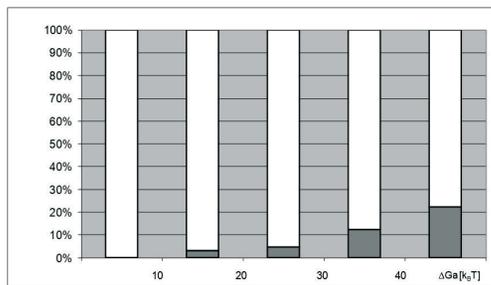
Figure 6. Fraction of results for large product molecule (DNA, RNA, proteins) among all results divided into five classes of activation energy.
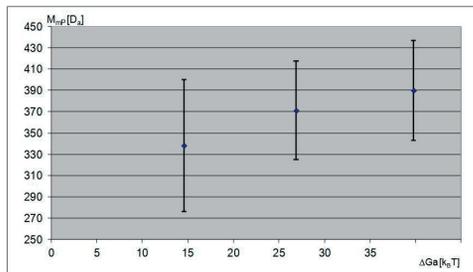
Figure 7. Mean product mass for different activation energies.
Calculated for 20 of the smallest energies, 20 of the highest energies, and random 20 energies in the range $\Delta G_a = 26–28$ [k_BT].
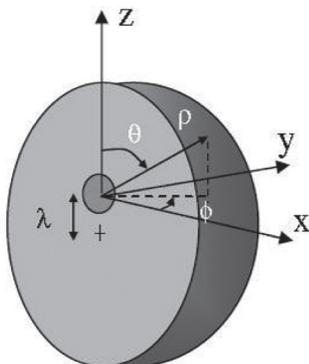
Figure 8. System of coordinates x,y,z and related spherical variables σ, θ, φ
(see Appendix 1 at http://www.actabp.pl/pdf/2_2013/239-suppl.pdf).

test the above, we looked for reactions with large product molecules (DNA, RNA, proteins) among the results of our activation energy calculations. Five classes of activation energy, $\Delta G_a \ 0 \leq 10$ [k_BT], …, $\Delta G_a > 40$ [k_BT], were considered. As it is seen in Fig. 6, for extremely small energies we observed no big products, but for extremely high energies — over of reactions were those of big molecules.

What about smaller molecules? To simply answer this question we calculated the mean product mass only for selected characteristic conditions: 20 smallest energies, 20 highest energies, and random 20 in the range $\Delta G_a =$ 26–28 [k_BT]. Results with DNA, RNA or proteins were excluded. For reactions with several products, the highest mass was taken for calculation. Results for the same EC and RE, but different MM, were taken once. The mean energy in a given group was calculated (Fig. 7).

In agreement with our model's prediction, there was a tendency for reactions with very low activation energies to have smaller products. Unfortunately, the scatter of data is very large and the correlation is not statistically

significant. Nevertheless, both above calculations tend to confirm the correctness of the proposed model.

CONCLUSIONS

Highlights

- A general dependence of the enzyme catalytic rate on its mass was demonstrated.
- The activation energy decreases with the enzyme mass increase.
- A theoretical model for the enzyme and departing product has been proposed.

Summing up, it was shown that statistical investigations using large sets of data from metabolic pathway databases can be useful in testing general ideas. In the presented case, basing on the observed mass–activation energy dependence, a new concept of the enzyme turnover number was formulated as a measure of the frequency of product release. According to the proposed theoretical model, the enzyme-product binding was attributed to local as well as global interactions. A collective of local hydrogen bonds and surface van der Waals forces in the catalytic site, and also global London interactions in the entire complex, were reasonably identified.

Despite some unanswered questions left and new areas waiting for future statistical exploration, e.g., the substrate–enzyme complex approaching the catalytic process, the presented paper can be viewed as the first attempt at physical generalization of enzymatic phenomena based on a large set of experimental data.

REFERENCES

Adams JA, Taylor JA (1992) Energetic limits of phosphotransfer in the catalytic subunit of cAMP-dependent protein kinase as measured by viscosity experiments. *Biochemistry* **31**: 8516–8522.

Aubol BE, Ungs L, Lukasiewicz R, Ghosh G, Adams JA (2004) Chemical clamping allows for efficient phosphorylation of the RNA carrier protein Npl3. *J Biol Chem* **279**: 30182–30188.

Alberty RA (1959) In *The Enzymes*. 2nd ed, Vol **1**: 130–150.

Antoniou D, Schwartz SD (1997) Large KIE in enzymatic proton transfer and the role of substrate oscillations. *Proc Natl Acad Sci* **94**: 12360–12365.

Bairoch A (2000) The ENZYME database in 2000. *Nucleic Acids Res* **28**: 304–305.

Bolam DN, Ciruela A, McQueen-Mason S, Simpson P, Williamson MP, Rixon JE, Borastons A, Hazlewood GP, Gilbert HJ (1998) Pseudomonas cellulose-binding domains mediate their effects by increasing enzyme substrate proximity. *Biochem J* **331**: 775–781.

Britt BM (1993) A shifting specificity model for enzyme catalysis. *J Theor Biol* **164**: 181–90.

Britt BM (1997) For enzymes, bigger is better. *Biophys Chem* **69**: 63–70.

Burgard AP and Maranas CD (2001) Review of the Enzymes and Metabolic Pathways (EMP) Database. *Metabolic Engineering* **3**: 193–194.

Cannon WR, Benkovic SJ (1998) Solvation, reorganization energy, and biological catalysis. *J Biol Chem* **273**: 26257–26260.

Chaplin MF, Bucke C (1990) In *Enzyme Technology*. pp 1–39. CUP Archive.

Cleland WW, Frey PA, Gerlt JA (1998) The low barrier hydrogen bond in enzymatic catalysis. *J Bio Chem* **273**: 25529–25532.

Cleland WW (1970) in *The Enzymes*. 3rd edn. Vol **2**: 1–67.

Eyring H (1935) The activated complex and the absolute rate of chemical reactions. *Chem Rev* **17**: 65-77.

Fersht A (1999) In *Structure and mechanism in protein science: A guide to enzyme catalysis and protein folding*. W.H. Freeman and Company, New York.

Fischer E (1894) Einfluss der Configuration auf die Wirkung den Enzyme. *Ber Dtsch Chem Ges* **27**: 2985–2993.

Huml K (1999) Computer-aided modelling of enzyme – substrate interaction. *Materials Structure* **6**: 8–10.

Jencks WP (1975) Binding energy, specificity and enzymic catalysis: the Circe effect. *Adv Enzymol Relat Areas Mol Biol* **43**: 219–410.

Karp PD, Riley M, Saier M, Paulsen IT, Paley SM and Pellegrini-Toole A (2000) The EcoCyc and MetaCyc databases. *Nucleic Acids Res* **28**: 56–59.

Kohen A, Klinman PJ (1998) Enzyme catalysis: beyond classical paradigms. *Acc Chem Res* **31**: 397–404.

Koshland Jr DE (1958) Application of a theory of enzyme specificity to protein synthesis. *Proc Natl Acad Sci* **44**: 98–104.

Kraut J (1988) How do enzymes work? *Science* **242**: 533–540.

Leskovac V (2003) In *Comprehensive enzyme kinetics*. pp 37-45. Kluwer Acdemic Publishers, New York, Boston, Dordrecht, London, Moscow.

Lerch HP, Mikhailov AS, Hess B (2002) Conformational-relaxation models of single-enzyme kinetics. *Proc Natl Acad Sci* **99**: 15410–15415.

Levitt M, Perutz MF (1988) Aromatic rings act as hydrogen bond acceptors. *J Mol Biol* **201**: 751–754.

London F (1937) The general theory of molecular forces. *Trans. Faraday Soc* **33**: 8–26.

Mattey M, Simoes D, Brown A, Fan X (1998) Enzymes with a low molecular weight. *Acta Chim Slov* **45**: 45–57.

Mesecar AD, Stoddard BL, Koshland Jr DE (1997) Orbital steering in the catalytic power of enzymes: small structural changes with large catalytic consequences. *Science* **227**: 202–206.

Michaelis L, Menten M (1913) Die kinetik der invertinwirkung. *Biochemistry Zeitung* **49**: 333–369.

Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ (1998) Automated docking using a Lamarckian genetic algorithm and empirical binding free energy function. *J Comput Chem* **19**: 1639–1662.

Myers JK, Pace CN (1996) Hydrogen bonding stabilizes globular proteins *Biophys J* **71**: 2033–2039.

Narayana N, Cox S, Xuong NH, Eyck LFT, Taylor SS. (1997) A binary complex of the catalytic subunit of cAMP-dependent protein kinase and adenosine further defines conformational flexibility, *Structure* **5**: 921–935.

Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M (1999) KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **27**: 27 29–34.

Pauling L (1946) Molecular architecture and biological reactions. *Chem Eng News* **24**: 1375–1377.

Pauling L (1960) In *The Nature of the Chemical Bond*. 3rd edn. Cornell Univ. Press. Ithaca, NY.

Plowman KM (1972) In *Enzyme Kinetics*. McGraw-Hill. New York.

San Martín MC, Stamford NPJ, Dammerova N, Dixon NE, Carazo JM (1995) A structural model for the Escherichia coli DnaB helicase based on electron microscopy data. *J Structural Biol* **11**: 167–176.

Schleinkofer K, Sudarko, Winn PJ, Lüdemann SK, Wade RC (2005) Do mammalian cytochrome P450s show multiple ligand access pathways and ligand channelling? *EMBO Reports* **6**: 584–589.

Schomburg I, Chang A, Schomburg D (2002) BRENDA, enzyme data and metabolic information. *Nucleic Acids Res* **30**: 47–49.

Segel IH (1993) Steady state kinetics of multireactant enzymes. In *Enzyme Kinetics*. Chapter Nine pp 506–543. Wiley. New York.

Selkov E, Basmanova S, Gaasterland T, Goryanin I, Gretchkin Y, Maltsev N, Nenashev V, Overbeek R, Panyushkina E, Pronevitch L, Selkov Jr E, Yunus I (1996) The metabolic pathway collection from EMP: the enzymes and metabolic pathways database. *Nucleic Acids Res* **24**: 26–28.

Sierks M. R,. Sico C, Zaw M. (1997) Solvent and viscosity effects on the rate-limiting product release step of glucoamylase during maltose hydrolysis. *Biotechnol Progress* **13**: 601–608.

Stoddard BL, Cohen BE, Brubaker M, Mesecar AD, Koshland Jr DE, (1998) Millisecond Laue structures of an enzyme-product complex using photocaged substrate analogs. *Nat Struct Biol* **5**: 891–897.

Tian G, Sobotka-Briner CD, Zysk J, Liu X, Birr C, Sylvester MA, Edwards PD, Scott CD and Greenberg BG (2002) Linear non-competitive inhibition of solubilized human γ-secretase by pepstatin a methylester, L685458, sulfonamides, and benzodiazepines. *J Biol Chem* **277**: 31499–31505.

Warshel A (1998) Electrostatic origin of the catalytic power of enzymes and the role of preorganized active Sites. *J Biol Chem* **273**: 27035–27038.

Willemsen OH, Snel MME, Cambi A, Greve J, de Grooth BG and Figdor CG (2000) Biomolecular interactions measured by atomic force microscopy. *Biophys J* **79**: 3267–3281.

Yin J, Andryski SE, Beuscher IV AE, Stevens RC, Schultz PG (2003) Structural evidence for substrate strain in antibody catalysis. *Proc Natl Acad Sci* **100**: 856–861.