# QUANTITATIVE ESTIMATION OF ACTIVITIES OF MUTANT ENZYMES

Gábor NÁRAY-SZABÓ

*CHINOIN Pharmaceutical and Chemical Works, P.O. Box 110, H-1325 Budapest, Hungary*

Received 6 October 1988; accepted 1 February 1989

A simple equation is presented for the prediction of catalytic efficiencies and Michaelis constants for pairs of mutant serine proteases and substrates with variable $P_1$ side chains. The equation is a generalization of the "similis simili gaudet" principle formulated by Gráf et al. [3] (Proc. Natl. Acad. Sci. U.S.A. 85 (1988) 4961, stating that in water amino-acid side chains with similar polarities tend to interact stronger than dissimilar ones.

Owing to the spectacular development in protein engineering a wealth of information is now available on the activities of mutant enzymes [1,2]. In a recent experimental paper an interesting trend has been formulated [3]. It was found that for the Ser/Asp-189 mutant of trypsin the tetrapeptide Succ-Ala-Ala-Pro-Lys-7-amino-4-methyl-coumarin is a better substrate if at position $P_1$ the positively charged Lys side chain is deprotonated, i.e. the pH is raised from 7 to 11. In order to maximize the catalytic efficiency ($\log k_{cat}/K_m$) the interacting side chains of the enzyme and its substrate have to be of similar polarity. Accordingly, side chain combinations where both are (oppositely) charged ($Asp^-$ or $Glu^-$- vs. $Lys^+$ or $Arg^+$), polar (protonated Asp and Glu, deprotonated Lys and Arg, Asn, Ser, Gln, Cys, Thr or His vs. each other) allow larger catalytic efficiences than pairs with electrostatically dissimilar side chains. The hypothesis was tested [4] using experimental data for trypsin [3], subtilisin [5,6] and tyrosyl tRNA-synthetase [7] and it was found to be valid for a considerable number of cases with a few outliers. We report here on rationalisation and refinement of this rule resulting in a simple phenomenological equation for the prediction of $\log k_{cat}/K_m$ and $\log K_m$ for pairs of mutant enzymes and substrates.

In the following we try to find a theoretical explanation for the "similis simili gaudet" principle. This will not be a proof, rather a heuristic foundation of the mathematical form of the quantitative structure-activity relationship presented later.

First of all, let us note that the catalytic efficiency depends linearly on the association free energy change, $\Delta G^\dagger$, of the enzyme-substrate complex in the transition state as compared to the ground state [8]. Mathematically, $\log k_{cat}/K_m = -\Delta G^\dagger/2.3RT$. $R$ is the gas constant and $T$ is the absolute temperature.

To estimate $\Delta G^{\dagger}$ let us consider the following thermodynamic process

$$(E)_{aq} + (S)_{aq} \xrightarrow{\Delta G_d} (E)_d + (S)_d \xrightarrow{\Delta G_{ass}} (E \cdot S)_{aq} \rightarrow \xrightarrow{\Delta G_{act}} (ES^{\dagger})_{aq} \quad (1)$$

where (E), (S), (E · S) and ($ES^{\dagger}$) denote the enzyme, the substrate, the Michaelis complex and the enzyme-substrate complex in the transition state, respectively. The subscript "aq" denotes the fully hydrated species, while "d" refer to a partially dehydrated state where active site and involved substrate side chain(s) are free for association to form the Michaelis complex. $\Delta G_{ass}$ and $\Delta G_{act}$ stand for association and activation free energies. $\Delta G^{\dagger}$ is now written as

$$\Delta G^{\dagger} = \Delta G_d + \Delta G_{ass} + \Delta G_{act} \quad (2)$$

In this paper we restrict to a series of mutant serine proteases and their substrates where side chains are varied exclusively in the specificity pocket and/or at the substrate $P_1$ site. Accordingly, we approximate the dehydration free energy as a sum of side-chain terms

$$\Delta G_d \approx \sum_i \Delta G_d(s_i^E) + \Delta G_d(P_1) \quad (3)$$

where $s_i^E$ refers to the $i$-th side chain in the specificity pocket interacting with the substrate $P_1$ side chain.

Let us suppose that the sum in eq. (3) includes only a single term, furthermore that dehydration free energies are independent of solvent reorganization and entropy effects and can be approximated by pure electrostatic energy terms, $\Delta G_d(s) \approx \Delta E_d(s)$ with $s = s^E$ or $P_1$. Based on the studies of Pullman et al. [9] we postulate that $\Delta E_d(s)$ is proportional to the average value of the electrostatic field, $F$, produced by the side chain on its van der Waals surface

$$\Delta E_d(s) = A \cdot F(s) + B \quad (4)$$

Furthermore, we assume that $\Delta G_{ass}$ is independent of entropy, steric and reorganization effects and can be approximated by an electrostatic term, $\Delta G_{ass} \approx \Delta E_{ass}$. We estimate $\Delta E_{ass}$ as a product of electrostatic potentials in a reference point near the interacting side chains as proposed by Kollman [10]

$$\Delta E_{ass} \approx C \cdot V(s^E) \cdot V(P_1). \quad (5)$$

If $V$ is purely Coulombic, i.e. $V \approx Q/r$ where $Q$ is an effective charge, and $r$ is the distance of the reference point from the charge centre,

$$F = \frac{d}{dr} V = -Q/r^2,$$

i.e. $V$ is proportional to the square root of the absolute value of $F$, $V \propto F^{1/2}$. On this basis we write heuristically

$$\Delta E_{ass} = C \cdot F^{1/2}(s^E) \cdot F^{1/2}(P_1). \quad (6)$$

Table 1
Average electrostatic fields on the van der Waals surface of amino-acid side chains (scale shifted by 10 V/nm)

| side chain | $F$(V/nm) | side chain | $F$(V/nm) |
|---|---|---|---|
| *charged* | | | |
| $Asp^-$ | 36.7 | His | 19.0 |
| $Glu^-$ | 32.8 | Lys (neutral) | 17.3 |
| Tyr | 15.0 | *apolar* | |
| $Lys^+$ | 28.1 | Met | 15.6 |
| $Arg^+$ | 25.8 | Tyr | 15.0 |
| *polar* | | Trp | 14.4 |
| Asn | 22.7 | Ala | 11.3 |
| Ser | 21.8 | Val | 11.3 |
| Arg (neutral) | 21.8 | Leu | 11.3 |
| Gln | 20.9 | Ile | 11.2 |
| Cys | 20.2 | Phe | 10.6 |
| Thr | 19.7 | Gly | 10.0 |

At last, we suppose that mutated and $P_1$ substrate side chains lie far enough from the tetrahedral intermediate formed in the transition state therefore do not interact with it, so $\Delta G_{act}$ is independent of mutations. Combining eqs. (2–6), considering that $\Delta G^{\dagger} \propto \log K_{cat}/K_m$ and putting $C/A = -2$ we have

$$\log k_{cat}/K_m = A \cdot X + \text{const.} \tag{7}$$

with

$$X = F(s^E) + F(P_1) - 2F^{1/2}(s^E) \cdot F^{1/2}(P_1). \tag{8}$$

We calculated $F$ values for amino-acid side chains using standard geometries and atomic point charges by Weiner et al. [11]. We considered molecular van der Waals surfaces in the Connolly dot surface representation [12]. Side chains were modeled by neglecting the peptide group of the residue. Electrostatic fields were calculated by Coulomb's law in 200 points of the surface that were averaged to obtain $F$.

Table 1 displays a shifted scale of amino-acid side chain fields with $F' = F + 10$. Using $F'$ in eq. (8) to calculate $X$ we compared $\log k_{cat}/K_m$ to $X$ in table 2 for a number of $s^E$/Gly-166 subtilisin mutants vs. various substrates. For Met and Lys substrates we dropped the Met/Gly-166 mutant and for all substrates the Arg/Gly-166 and Lys/Gly-166 mutants from the data set since it is highly probable that for these pairs of interacting side chains steric effects cannot be neglected as supposed in the rationalization of eq. (5). The constant in eq. (7) is identical for the isosteric side chains Met and Lys while it is shifted by 3 V/nm for the $Glu^-$, Gln pair. This transformation represents the difference of the constant $B$ of eq. (4) for side chains of different size.

Table 2
Catalytic efficiencies and $X$ values of eq. (8) for Gly/$s^E$-166 mutants of substilisin vs. various substrates. Experimental data are from ref. [6]

| $s^E$ | Met | | $Lys^+$ | |
|---|---|---|---|---|
| substrate | log $k_{cat}/K_m$ | $X$ | log $k_{cat}/K_m$ | $X$ |
| Asp | 3.81 | 4.45 | 4.21 | 0.57 |
| Glu | 3.86 | 3.16 | 4.48 | 0.18 |
| Asn | 5.02 | 0.66 | 4.25 | 0.29 |
| Gln | 5.54 | 0.39 | 4.10 | 0.53 |
| Ala | 5.65 | 0.35 | 4.90 | 3.76 |
| Gly | 5.15 | 0.62 | 4.60 | 4.57 |
| $s^E$ | $Glu^-$ | | Gln | |
| substrate | log $k_{cat}/K_m$ | $X+3$ | log $k_{cat}/K_m$ | $X+3$ |
| Asp | – | 3.11 | 3.02 | 5.21 |
| Glu | – | 3.00 | 3.06 | 4.34 |
| Asn | 1.62 | 3.93 | 3.85 | 3.04 |
| Gln | 1.20 | 4.34 | 4.36 | 3.00 |
| Met | 1.20 | 6.16 | 3.89 | 3.39 |
| Ala | – | 8.60 | 4.34 | 4.46 |
| Gly | 1.54 | 9.58 | 3.95 | 4.99 |

log $k_{cat}/K_m$ is plotted vs. $X$ in fig. 1. It is seen that except three cases with Glu as the $P_1$ side chain in the substrate the points lie close to a decreasing straight line reaching its maximum at $X = 0$, i.e. at $F'(s) = F'(P_1)$. So eq. (7) is a semiquantitative counterpart of the "similis simili gaudet" principle formulated quantitatively in ref. [3].

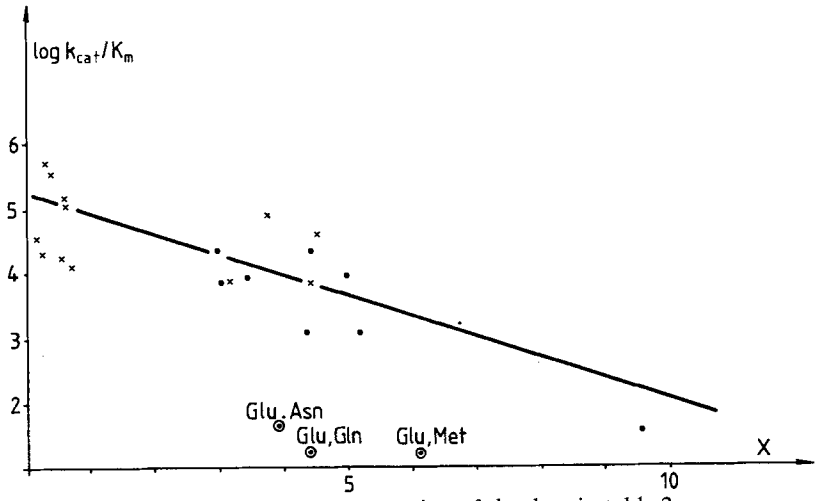

Fig. 1. Graphic representation of the data in table 2.

Table 3
Statistical parameters for eq. (9) $n$, $r$ and $s$ are the number of data points, correlation coefficient and standard error, respectively.

| case | $a$ | $b$ | $c$ | $d$ | $n$ | $r$ | $s$ |
|---|---|---|---|---|---|---|---|
| subtilisin [6] $1/K_m$ | −0.049 | −0.100 | −0.013 | 7.06 | 47 | 0.907 | 0.32 |
| subtilisin [6] $k_{cat}/K_m$ | −0.073 | −0.182 | −0.038 | 10.49 | 47 | 0.964 | 0.35 |
| trypsin [3] $K_m$ | 0.011 | 0.036 | −0.007 | 2.48 | 7 | 0.912 | 0.30 |
| trypsin [3] $k_{cat}/K_m$ | −0.003 | 0.044 | −0.088 | 2.27 | 7 | 0.993 | 0.40 |

Equation (7) can be generalized as follows

$$\log k_{cat}/K_m = a \cdot F(s^E) + b \cdot F(P_1) + c \cdot q \cdot F^{1/2}(s^E) \cdot F^{1/2}(P_1) + d. \quad (9)$$

To distinguish between electrostatic attraction and repulsion between charged side chains, heuristically we put $q = -1$ if both $s^E$ and $P_1$ are oppositely charged and $q = +1$ otherwise. Dropping sterically interacting side-chain combinations from the data set as above, eq. (9) gives good correlation not only for subtilisin [6] but also for trypsin (cf. table 3). The same type of equation holds for log $K_m$ but the correlation is somewhat worse. This can be rationalized on the basis of eqs. (1–2) considering that $\log K_m \propto \Delta G_d + \Delta G_{ass}$.

Equation (9) seems to be appropriate to estimate catalytic efficiencies and Michaelis constants for various pairs of enzymes mutated at their specificity pockets and substrates substituted at site $P_1$ prior to experiment or more sophisticated theoretical calculations [13–15].

## Acknowledgements

I am indebted to Prof. L. Gráf (Budapest) for illuminating discussions and to Mr. I. Kolossváry (Budapest) for writing the computer program for the calculation of average electrostatic fields of amino-acid side chains.

## References

[1] J.R. Knowles, Science 236 (1987) 1252–1258.
[2] W.V. Shaw, Biochem. J. 246 (1987) 1–17.
[3] L. Gráf, Á. Jancsó, G. Hegyi, K. Pintér, G. Náray-Szabó, J. Hepp, K. Medzihradszky and W.J. Rutter, Proc. Natl. Acad. Sci. U.S.A. 85 (1988) 4961–4965.
[4] G. Náray-Szabó, J. Mol. Catal. 47 (1988) 281–287.
[5] D.A. Estell, T.P. Graycar, J.V. Miller, D.B. Powers, J.P. Burnier, P.G. Ng and J.A. Wells, Science 233 (1986) 659–663.
[6] J.A. Wells, D.B. Powers, R.R. Bott, T.P. Graycar and D.A. Estell, Proc. Natl. Acad. Sci. U.S.A. 84 (1987) 1219–1223.

[7] A.R. Fersht, J.P. Shi, J. Knill-Jones, D.M. Lowe, A.J. Wilkinson, D.M. Blow, P. Brick, P. Carter, M.M.Y. Waye and G. Winter, Nature (London) 314 (1985) 235–238.
[8] A.R. Fersht, *Enzyme Structure and Mechanism*, 2nd ed. (Freeman, San Francisco, 1985).
[9] A. Pullman, B. Pullman and R. Lavery, J. Mol. Struct. THEOCHEM 93 (1983) 85–91.
[10] P. Kollman, J. Am. Chem. Soc. 99 (1977) 4875–4894.
[11] S.J. Weiner, P.A. Kollman, D.A. Case, U.C. Singh, C. Ghio, G. Alagona, S. Profeta and P. Weiner, J. Am. Chem. Soc. 106 (1984) 765–775.
[12] M.L. Connolly, Science 221 (1983) 709–713.
[13] A. Warshel and F. Sussman, Proc. Natl. Acad. Sci. U.S.A. 83 (1986) 3806–3810.
[14] M.K. Gilson and B.H. Honig, Nature (London) 330 (1987) 84–86.
[15] M.J.E. Sternberg, F.R.F. Hayes, A.J. Russell, P.G. Thomas and A.R. Fersht, Nature (London) 330 (1987) 86–88.