THE QUANTIFICATION OF PROTEIN BIOACTIVITY. PHAGE T4 LYSOZYMES SUBSTITUTED AT RESIDUE 86

Marvin CHARTON

Chemistry Department, School of Liberal Arts and Sciences, Pratt Institute, Brooklyn, New York 11205, U.S.A.

Received February 14, 1989
Accepted March 24, 1989

Dedicated to Professor Otto Exner on the occasion of his 65th birthday.

Data taken from the literature for the activity of mutant Phage T4 lysozymes substituted at residue 86 were correlated with the intermolecular force (IMF) equation with very good results. The activity determined was the relative amount of enzyme required to give a cleared area with a radius of 7·00 mm in the lysoplate assay of Becktel and Baase. The best regression equation obtained shows hydrogen bonding to be the major factor in determining activity, with dispersion forces next in importance and steric effects least. No dependence on side chain charge was observed. The results support the utility of the IMF equation as a quantitative model of protein bioactivity.

It has seemed to us to be of great interest and importance to determine quantitatively the effect of amino acid residue side chain structure on the biological activities of proteins. In order to carry out such studies it is necessary to have available data for the effect on bioactivity of the replacement of one residue by another in protein molecules. Ideally, all of the replacements should be at the same position in the protein and all of the amino acid residues commonly found in protein should be included in the data set. There are unfortunately very few publications extant in the literature which have reported the effect of substitution on protein bioactivity. We have previously shown that the values of $\Delta\Delta G$ determined by kinetic methods for transition state binding of ATP and tyrosine by tyrosyl-t-RNA synthetase by Fersht and coworkers[1] are quantitatively described by the intermolecular forces (IMF) equation[2]. Alber and coworkers[3] have recently reported relative activities of Phage T4 lysozymes substituted at residue 86. The activity reported is the relative amount of enzyme required to give a cleared area with a radius of 7·0 mm in the lysoplate assay of Becktel and Baase[4]. In this work we have applied the IMF equation to the quantitative description of the relative bioactivities of these Phage T4 lysozymes substituted at residue 86. Our objective is twofold: to determine if possible the nature of the dependence of the activity of these enzymes on side chain structure, and to provide a further test of the utility of the IMF equation in the quantitative description of protein bioactivity.

THEORETICAL

The IMF *Equation*

The IMF equation is based on the assumption that the magnitude of a property or activity is determined by the change in intermolecular forces resulting from the transfer of a substrate from an initial to a final phase. Examples of such transfers are from one solvent to another, from the solid state to a solution, from a solvent to the surface of a chromatographic column, from an aqueous phase to a biomembrane surface, from an aqueous phase to the active site of an enzyme, and from one protein conformation to another. The IMF equations has been successfully applied to the quantitative description of amino acid hydrophobicities, transport properties and bioactivities[5-8], peptide transport properties[9] and bioactivities[10,11], and protein bioactivities[2] and conformational parameters[12-14]. It may be written in its most general form as either Eq. (*1*)

$$Qx = L\sigma_{1X} + D\sigma_{dX} + R\sigma_{eX} + A\alpha_X + H_1 n_{HX} + H_2 n_{nX} + Ii_X + C_A n_{AX} + \\ + C_D n_{DX} + Sv_X + B^0 \qquad (1)$$

or Eq. (*2*)

$$Q_X = L\sigma_{1X} + D\sigma_{dX} + R\sigma_{eX} + A\alpha_X + H_1 n_{HX} + H_2 n_{nX} + Ii_X + C_A n_{AX} + \\ + C_D n_{DX} + B_1 n_{1X} + B_2 n_{2X} + B_3 n_{3X} + B^0 . \qquad (2)$$

depending on the choice of steric effect parameterization. Qx represents the quantity which is to be corelated. σ_1 σ_p and σ_e are the localized (field and/or inductive), intrinsic delocalized (resonance), and electronic demand sensitivity electrical effect parameters respectively[13]. The σ_1 parameters are identical to the σ_I parameters. The electrical effect parameters are involved in modelling in whole or in part contributions from dipole–dipole (dd), dipole induced dipole (di), hydrogen bonding (hb), charge transfer (ct), and ion-dipole (Id) interactions. α is a polarizability parameter derived from group molar refractivities, it models dispersion forces (induced dipole–induced dipole, ii, interactions) and to some extent di interactions and ion induced dipole (Ii) contributions as well. The parameters n_H and n_n model hydrogen bonding; the former is defined as the number of OH and/or NH bonds in the X group while the latter is defined as the number of lone pairs on N and/or O atoms in that group. The *i* parameter represents the charge on the group X, it is defined as 1 for ionic X and 0 for nonionic X. It is involved in modelling the Id and Ii contributions. The n_A and n_D parameters are involved in modelling the ct contribution. The former is defined as 1 for an X group which can act as a charge transfer acceptor and 0 for one which cannot, the latter is similarily defined as 1 for an X group which can act as a charge transfer donor and 0 for one which cannot.

The υ parameter is a steric parameter defined from the Van der Waals radii[15]. Its merit is its simplicity, its disadvantage is that for most groups no single steric parameter can describe their steric effects in all systems. In many cases it is necessary to use the branching parameters n_i. These topological steric parameters are defined as the number of atoms other than H bonded to the i-th atoms of the X group[16]. We have frequently found them necessary for modelling steric effects in peptide and protein properties and bioactivities.

Correlations

We have found that in the correlation of the properties and bioactivities of amino acids, peptides, and proteins only the localized electrical effect parameter is required. Furthermore, we have generally failed to find a dependence on the charge tranfer parameters n_A and n_D. Finally, the small size of the data set to be studied precludes the use of branching parameters to model the steric effect. We have therefore used the correlation Eq. (*3*)

$$\log A_{rel,X} = L\sigma_{1X} + A\alpha_X + H_1 n_{HX} + H_2 n_{nX} + Ii_X + S\upsilon_X + B^0 . \qquad (3)$$

The data were correlated with Eq. (*3*) by multiple linear regression analysis. The data used in the correlations are set forth in Table I together with the parameters used.

RESULTS

Correlation with Eq. (*3*) gave the regression Eq. (*4*)

$$\log A_{rel,X} = 0{\cdot}373\,(\pm 0{\cdot}632)\,\sigma_{1X} + 2{\cdot}09\,(\pm 1{\cdot}19)\,\alpha_X - \\ - 0{\cdot}319\,(\pm 0{\cdot}0771)\,n_{HX} + 0{\cdot}174\,(\pm 0{\cdot}0586)\,n_{nX} - 0{\cdot}0889\,(\pm 0{\cdot}115)\,i_X - \\ - 0{\cdot}463\,(\pm 0{\cdot}245)\,\upsilon_X - 0{\cdot}148\,(\pm 0{\cdot}0701) \qquad (4)$$

$$100R^2 = 94{\cdot}59\,; \quad F = 8{\cdot}749\,; \quad s_{est} = 0{\cdot}0790\,; \quad s_0 = 0{\cdot}424\,; \quad n = 10\,.$$

The correlation matrix is given in Table II. Both L, the coefficient of σ_1X, and I, the coefficient of i_X are not significant. On repeating the correlation with the exclusion of the σ_1X parameter we obtained the regression equation

$$\log A_{rel,X} = 2{\cdot}24\,(\pm 1{\cdot}07)\,\alpha_X - 0{\cdot}337\,(\pm 0{\cdot}0649)\,n_{HX} + \\ + 0{\cdot}192\,(\pm 0{\cdot}0461)\,n_{nX} - 0{\cdot}0860\,(\pm 0{\cdot}105)\,i_X - 0{\cdot}495\,(\pm 0{\cdot}219)\,\upsilon_X - \\ - 0{\cdot}135\,(\pm 0{\cdot}0613) \qquad (5)$$

$$100R^2 = 93{\cdot}97\,; \quad F = 12{\cdot}46\,; \quad s_{est} = 0{\cdot}0723\,- \quad s_0 = 0{\cdot}388\,; \quad n = 10\,.$$

TABLE I

Data and parameters used in the correlations[a]

Aax	A_{rel}	σ_1	α	n_H	n_n	i	υ
Pro[b]	1·0	−0·01	0·140	0	0	0	−0·81
Ala	0·5	−0·01	0·046	0	0	0	0·52
Arg	0·2	0·04	0·291	4	3	1	0·68
Asp	1·1	0·15	0·105	1	4	1	0·76
Cys	0·7	0·12	0·128	0	0	0	0·62
Gly	0·7	0	0	0	0	0	0
His	0·7	0·08	0·230	1	1	1	0·70
Ile	0·7	−0·01	0·186	0	0	0	1·02
Leu	0·5	−0·01	0·186	0	0	0	0·98
Ser	0·7	0·11	0·062	1	2	0	0·53
Thr	0·8	0·04	0·108	1	2	0	0·50
Asn	—	0·06	0·134	2	3	0	0·76
Glu	—	0·07	0·151	1	4	1	0·68
Gln	—	0·05	0·180	2	3	0	0·68
Lys	—	0·00	0·219	2	1	1	0·68
Met	—	0·04	0·221	0	0	0	0·78
Phe	—	0·03	0·290	0	0	0	0·70
Trp	—	0·00	0·409	1	0	0	0·70
Tyr	—	0·03	0·298	1	2	0	0·70
Val	—	0·01	0·140	0	0	0	0·76

[a] Data are from ref.[3], parameter values are from refs[5,6]. Aax are residues in position 86. A_{rel} are the relative activities determined as described in ref.[4], the substrate is *E. coli* peptidoglycan; [b] wild enzyme; this data point was not included in the correlation.

TABLE II

Correlation matrix

σ_1	α	n_H	n_n	i	υ	
1	0·007	0·197	0·608	0·445	0·019	σ_1
	1	0·603	0·213	0·576	0·671	α
		1	0·698	0·674	0·051	n_H
			1	0·686	0·071	n_n
				1	0·199	i
					1	υ

As I is still not significant the correlation was repeated with the exclusion of the parameter *i*, giving Eq. (*6*)

$$\log A_{\mathrm{rel,X}} = 1{\cdot}63\,(\pm 0{\cdot}747)\,\alpha_{\mathrm{X}} - 0{\cdot}311\,(\pm 0{\cdot}0548)\,n_{\mathrm{HX}} + \\ + 0{\cdot}164\,(\pm 0{\cdot}0300)\,n_{\mathrm{nX}} - 0{\cdot}363\,(\pm 0{\cdot}174)\,\upsilon_{\mathrm{X}} - 0{\cdot}132\,(\pm 0{\cdot}0591) \qquad (6)$$
$$100R^2 = 92{\cdot}95\,; \quad F = 16{\cdot}48\,; \quad s_{\mathrm{est}} = 0{\cdot}0699\,; \quad s_0 = 0{\cdot}376\,; \quad n = 10\,.$$

The best equation obtained, Eq. (*6*), is significant at the 99·5% confidence level. It must be pointed out that the possibility of an accidental correlation must always be considered, even more so when as is the case here the number of degrees of freedom (DF) is so small. In this data set there are only 10 data points, as a result in Eq. (*6*) there are only 1·25 DF per independent variable. Had we attempted to model this data set by choosing at random from the very large number of available physicochemical parameters until we finally arrived at some that seemed to give a good fit we would not be entitled to draw any conclusions from our results. We have not done this however. The IMF equation has been applied successfully to more than two hundred sets of amino acid, peptide, and protein properties and bioactivities using the same parameter set. Certainly the possibility of accidental correlation remains, but the probability of more than two hundred accidental correlations is not merely remote, it is asymptotic with zero. We believe therefore that Eq. (*6*) is probably meaningful and does provide a quantitative description of the effect of side chain substitution on the measured bioactivity.

The effect of the side chain of the residue in position 86 on the activity is due to dispersion forces, hydrogen bonding and to steric effects. The contribution made by each variable is quantitatively described by the quantity C_i which is defined by the relationship (*7*)

$$C_i = |b_i x_i| \cdot 100 / \sum_{i=1}^{n} |b_i x_i| \,, \qquad (7)$$

where the b_i are the coefficients of the independent variables in the IMF equation and the x_i are the values of these variables for some reference side chain X^0. In our applications of the IMF equation to properties of amino acids, peptides and proteins we have chosen the side chain of His as the reference side chain. The values of C_i obtained for Eq. (*6*) are: α, 33·9; n_{H}, 27·6; n_{n}, 14·5; and υ, 24·4. Thus the total hydrogen bonding contribution of 42·1 is the largest, followed by dispersion forces, with steric effects making the smallest contribution.

We define the supplier of the lone pair in a hydrogen bond as the donor and the source of the hydrogen atom as the acceptor. The sign of H_1, the coefficient of n_{H}, is negative showing that acceptor activity in the initial phase is larger than in the final phase. By contrast, the sign of H_2, the coefficient of n_{n}, is positive indicating that there is more donor activity in the final than in the initial phase. The positive

sign of A, the coefficient of α shows that dispersion forces are producing a larger contribution in the final than in the initial phase, suggesting that the side chain of residue 86 is making more and/or shorter contacts in the final phase than in the initial one. Finally the negative sign of S, the coefficient of the steric parameter υ, shows that steric repulsion is greater in the final phase than it is in the initial. We may conclude then that in the final phase the side chain in position 86 is making more contacts with lone pairs on O or N atoms and fewer with NH or OH bonds, and that it has moved closer either to some group of atoms in the substrate or to the side chain of a residue in some other part of the protein resulting in an increase in dispersion forces and in steric repulsion. The lack of dependence on the side chain charge parameter i or on the localized electrical effect parameter σ_1 may mean that there is no change in the magnitude of Id and Ii or of dd and di interactions on going from the initial to the final phase.

Calculated and observed values of the activities for the substituted lysozymes are reported in Table III together with the values of Δ (log A) calculated from the relationship (*8*).

TABLE III
Values of log A_{calc} and Δ (log A)

Aax	log A_{obs}	log A_{calc}	Δ (log A)	A_{calc}
Pro	0·000	0·415	−0·415	2·6
Ala	−0·301	−0·261	−0·040	
Arg	−0·699	−0·677	−0·022	
Asp	0·041	0·084	−0·043	
Cys	−0·155	−0·166	0·011	
Gly	−0·155	−0·132	−0·023	
His	−0·155	−0·179	0·024	
Ile	−0·155	−0·229	0·074	
Leu	−0·301	−0·213	−0·088	
Ser	−0·155	−0·223	0·068	
Thr	−0·097	−0·136	0·039	
Asn	—	−0·342	—	0·5
Glu	—	0·192	—	1·6
Gln	—	−0·267	—	0·5
Lys	—	−0·500	—	0·3
Met	—	−0·078	—	0·8
Phe	—	0·066	—	1·2
Trp	—	−0·051	—	0·9
Tyr	—	0·096	—	1·2
Val	—	−0·202	—	0·6

$$\Delta(\log A) = \log A_{obs} - \log A_{calc} \qquad (8)$$

In no case is the value of $\Delta(\log A)$ more than twice the standard error.

The residue occupying position 86 in the wild enzyme is Pro. The data point for this enzyme was not included in the correlation because we have shown[17] that the modelling of Pro and its derivatives requires an additional variable which accounts for substitution on the nitrogen atom of the peptide bond derived from the Pro type residue. To make the inclusion of this variable meaningful it is necessary to have at least two residues of the Pro type in the data set. We can calculate a value of log A for Pro from Eq. (6), however, and compare this with the experimentally observed value. If they agree then the effect of α-nitrogen substitution is negligible, if they differ significantly then the effect is important. The observed value is 0, the calculated value is 0·415. As the standard error of the estimate is 0·0699 a Student t test for the significance of the difference shows it to be significant at the 99·9% confidence level. It seems likely then that substitution at the α-nitrogen of the Pro residue which is not accounted for by its value of the steric parameter is significant.

This difference may be due to the loss of hydrogen bonding capability by the peptide nitrogen atom. Alternatively it may be due wholly or in part to the fact that the secondary structure of the wild enzyme differs from that of the mutants to a considerable extent in the region near residue 86 (ref.[3]).

DISCUSSION

Lysozyme hydrolyzes the β (1 → 4) glycosidic bond between N-acetylglucosamine and N-acetylmuramic acid in bacterial cell wall peptidoglycan. On the basis of X-ray crystallographic and chemical evidence[17,18] Asp 20 certainly is involved in the catalytic activity. Glu 11, which forms an ion pair with Arg 145 that is about 5 Å from the bond being cleaved, may also be involved. The other residues which are certainly involved in the catalytic activity are Gln 105 and Trp 138, while Glu 22, Asn 140 and Glu 141 are probably involved. The residue in position 86 is said to be about 24 Å from the bond undergoing scission and is on the surface of the enzyme molecule. The results we have obtained suggest that the effect of substitution at position 86 may be due to an effect on substrate binding to the enzyme, possibly the result of an effect on the population of a conformation which binds substrate to the enzyme in the transition state. The suggetion of Alber[3] that the decreased activity in the Arg substituted enzyme is due to the effect of the positive charge on the Arg side chain cannot be excluded but seems unlikely in view of our results. We find no dependence of activity on side chain charge. Furthermore, consideration of the values of the activity calculated for those amino acids which are commonly found in proteins but are not present in the data set (Table III) shows that Lys, for which the side chain must also be positive, gives a calculated value for the activity

of 0·3, comparable to that of Arg, though it was calculated from Eq. (*6*) which has no dependence on side chain charge. From the calculated values for mutant enzyme activity we see that the most active should be that in which Glu is substituted at position 86. From the calculated value of 1·6 obtained for the activity of this enzyme and the value of 0·2 determined experimentally for the Arg substituted enzyme, which is the least active, we see that substitution has at most an eightfold effect on enzyme activity. It would be helpful in determining why substitution at position 86 affects the activity if X-ray crystallographic studies of mutant enzyme-substrate complexes were available.

CONCLUSION

The IMF equation quantitatively describes the effect of substitution at residue 86 upon the activities of mutant Phage T4 lysozymes. Hydrogen bonding is the most important factor with dispersion forces second and steric effects third in importance.

In the final phase residue 86 apparently is making more contacts with lone pairs on O or N atoms and fewer with OH and/or NH bonds. It seems to be closer to either substrate or to some other residue in the enzyme, thus increasing the contribution of dispersion forces and concurrently increasing steric repulsion. This result supplies further support for the utility of the IMF equation in the quantification of protein bioactivities as a function of side chain structure.

REFERENCES

1. Fersht A. R., Shi J.-P., Knill-James J., Lowe D. M., Wilkinson A. J., Blow D. M., Brick P., Carter P., Waye M. M. Y., Winter G.: Nature *314*, 235 (1985).
2. Charton M.: Int. J. Pept. Protein Res. *28*, 201 (1986).
3. Alber T., Bell J. A., Dao-Pin S., Nicholson H., Wozniak J. A., Cook S., Matthews B. W.: Science *239*, 631 (1988).
4. Becktel W. J., Baase W. A.: Anal. Biochem. *150*, 258 (1985).
5. Charton M., Charton B. I.: J. Theor. Biol. *99*, 629 (1982).
6. Charton M., Charton B. I. in: *Quantitative Approaches to Drug Design* (J. Dearden, Ed.), p. 69. Elsevier, Amsterdam 1983.
7. Charton M., Charton B. I.: *Abstr. 3rd Noordwijkerhout Symposium* (*Innovative Approaches in Drug Research*), p. 102, (1985).
8. Charton M. in: *QSAR in Drug and Toxicology* (D. Hadzi and B. Jerman-Blazic, Eds), p. 291. Elsevier, Amsterdam 1987.
9. Charton M., Charton B. I.: *Abstr. 3rd Int. Conf. Correlation Analysis in Organic and Biological Chemistry* (Louvaine-la-Neuve), p. B16 (1985).
10. Charton M.: *QSAR and Strategies in the Design of Bioactive Compounds* (J. K. Seydel, Ed.), p. 260. Verlag Chemie, Weinheim 1985.
11. Charton M., Charton B. I. in ref.[8], p. 285.
12. Charton M., Charton B. I.: J. Theor. Biol. *102*, 121 (1983).
13. Charton M.: Abstr. 18th M.A.R.M. Am. Chem. Soc. (Newark, N.J.), p. 109, (1984).

14. Charton M., Charton B. I.: *Abstr. 3rd Noordwijkerhout Symposium* (*Innovative Approaches in Drug Research*) p. 101, (1985).
15. Charton M.: Top. Curr. Chem. *114*, 57 (1983).
16. Charton M.: Environ. Health Perspect. *61*, 229 (1985).
17. Charton M., Charton B. I. in ref.[6], p. 260.
18. Anderson W. F., Grutter M. G., Remington S. J., Weaver L. H., Matthews B. W.: J. Mol. Biol. *147*, 523 (1981).
19. Tsugita A. in: *The Enzymes* (P. D. Boyer, Ed.), p. 343. Academic Press, New York, 1971.