Optimizing model comparison for enzymatic mechanism analysis

AUTHOR NAMES:

Ryan Walsh^{†§}, Philippe Blain[‡]*

AUTHOR ADDRESS

† Armand-Frappier Santé Biotechnologie Research Centre, Institut national de la recherche scientifique, 531 boulevard des Prairies, Laval (Quebec), H7V 1B7, Canada.

§ Ronin Institute, 127 Haddon Pl, Montclair, New Jersey, 07043 USA.

ryan.walsh@ronininstitute.org

‡ Department of Chemistry, Université de Montréal, 2900 boulevard Édouard-Montpetit, Montréal (Quebec) H3C 3J7, Canada

KEYWORDS: Enzyme Kinetics, Inhibitors, Activators, Enzyme mechanism.

ABSTRACT

Characterization of enzyme inhibition in drug development is usually limited to a basic analysis with the classical inhibition models or simply the use of IC₅₀ values. However, a better understanding of enzyme physiology and regulation is seen as key to unraveling and treating the processes associated with stubborn disease targets like Alzheimer's disease. Recently it has been shown that enzyme regulation, through substrate, inhibitor or activator interactions can be modeled using the summation of binding curves. Here we examine the use of the modular equation permutations, that can be produced through binding curve summation, to fit and evaluate the interactions of abietic acid with protein tyrosine phosphatase nonreceptor type 11. This new sort of analysis will allow for improved insight into the physiological role enzymes play and the consequence their modulation may have in disease progression.

Introduction

Mechanistic analysis of enzyme systems is recognized as key to the understanding of physiology and disease processes (1). Current enzymatic analysis relies heavily on simplified models that work well with high throughput screening (1). This reliance may be due in part to the large number of models and equations which have been developed to describe complicated kinetic interactions. Recently it has been shown that the modulation of enzyme activity either inhibitory or stimulatory can be modeled using binding curves that equate changes in activity to the fractional association of the modifier with the enzyme population (2; 3). This results in distinct catalytic states that can be defined by the molecules associated with the enzyme (Figure 1). By equating changes in activity with binding curves it is possible to establish a modular expansion of the equations defining these binding curves to identify the catalytic states (4) (Figure 2). Here we take that concept one step further by introducing Hill coefficients to the binding curves improving the usefulness of these equations for modeling complex enzymatic mechanisms. This addition simplifies the creation of models of increasing complexity that can then be applied to datasets to evaluate the likelihood of potential interactions. This methodology is applied to the inhibition of protein tyrosine phosphatase nonreceptor type 11 by abietic acid (5). Protein tyrosine phosphatase nonreceptor type 11 is a member of a family of phosphatases that exhibit dynamic substrate interactions, in that they are known to rearrange when interacting with their larger peptide targets (6) and when interacting with small substrate analogues some of them are subject to substrate inhibition (7; 8). As this family of enzymes display complex regulation both by drug candidates (9) and their substrates the ability to evaluate multiple mechanisms clearly demonstrates the utility of this approach.

Materials and Methods

The previously published dataset for abietic acid inhibition of protein tyrosine phosphatase nonreceptor type 11 (E.C. 3.1.3.48; accession ID: Q06124) was used in this study (5). All non-linear regression model fitting was done using the solver feature of excel (10). Model development was based upon previously described methodologies (2; 4; 11). Models were evaluated based on their ability to fit the data using the sum of squared residuals (SSR) and Akaike information criterion (AIC) values (12).

Results and Discussion

In order to analyze the inhibition of SHP2 by AA, ten potential enzyme and inhibitor systems were examined. These systems focused on a maximum of two substrate and two inhibitor interactions with the enzyme (Figure 1; Table 1) resulting in a total of 204 potential inhibition models. As expected, the SSR decreased as the complexity of the models increased. This trend was very apparent with the one substrate one inhibitor system (Table 1a). The addition of Hill coefficients to the substrate-binding curve (Figure 3b; SSR 0.01247) or the inhibitor binding curve (Figure 3c; SSR 0.00899) were clear improvements over the basic equation (Figure 3a; SSR 0.01678). The addition of Hill coefficients to both curves further improved the fit when compared to the other three models (Figure 3d; SSR 0.00420). While in general, increases in model complexity improved the fit, within the two substrate systems a clear improvement in fitting was observed when Hill coefficients were applied to the first substrate-binding term. This

trend can be clearly seen in the SSR values associated with the fitting of models in section h where a green streak in column two is associated with the inclusion of a Hill coefficient in the first substrate-binding term but is absent in the third column which lists fits associated with Hill coefficients being added to the second substrate-binding term (Table 1h). This was not unexpected as it is known that SHP2 undergoes a rearrangement when interacting with peptide substrates (6; 13) but appears to experience substrate activation when examined with the small substrate analogue p-nitrophenyl phosphate used in this data set (11).

As the complexity of the models ranged between 5 and 22 parameters AIC values were calculated (Table 1) to discriminate between potentially useful models and those which were suffering from overfitting (12). The AIC values suggested that the single substrate model with two Hill coefficients (Figure 3d) potentially provided a useful model of the data (Table 1a). It also indicated that the addition of a second inhibitor to the single substrate system may provide a decent description of the data, with AIC values approaching the global minimum of the data set (Table 1b).

However, it has already been shown that based on plots of the noninhibited enzyme it is probably more appropriate to model the activity of this enzyme using a two-substrate binding model (11). Examining the two substrate models suggested several potentially useful models (Figure 4) associated with single inhibitor interactions with the substrate-induced forms of the enzyme (Table 1c-e). The AIC global minimum appeared with the models describing two inhibitors binding to the non-substrate activated form of the enzyme (Table 1f; Figure 4c). While more complex models improved on the fitting based on SSR values few produced viable models based on their increase in complexity (Figure 4d). Examining the actual binding constants for the model with the AIC minimum suggested a problem with the model, as for this model to achieve

the fit the secondary substrate binding constant had inflated to 1.7 M with an associated reaction rate 5.5 $\mu\text{M/s}$. Both of these values were quite high when contrasted to the substrate range used in the experiments which do not exceed 20 mM and the observed rates 0.28 $\mu\text{M/s}$. This large increase in rate can be observed as a slightly more pronounced upward swoop in the curve at higher substrate concentrations on the right side of the plot (Figure 4c). The same problem was observed with the simpler single inhibitor model (figure 4a), suggesting this effect was associated with models lacking inhibitor interactions with the substrate activated form of the enzyme. When models were allowed to include inhibitor interactions with the substrate activated form of the enzyme the substrate-binding term for the secondary binding were observed to fall much closer to or within the experimental dataset and changes in catalytic rate were generally doubled rather than suggesting up to twenty-fold increases. Models including inhibition of the substrate activated form also did not display as significant of an upward swoop when compared to models lacking this feature (Figure 4b, 4d). This suggests that while the AIC values of this dataset may favor models where the inhibitor only affects the non-substrate activated form of the enzyme, more appropriate models probably need to include inhibition of the substrate activated form of the enzyme.

Conclusions:

The depth and quantity of information provided by this sort of analysis is currently unavailable in drug and biological studies. While currently labor-intensive, the modular and systematic layout of the workflow involved in this study suggests that it will be amenable to automation and integration into modern bioinformatics approaches. Also, the ability to modularly construct

complex enzyme kinetic equations should greatly benefit research into disorders such as Alzheimer's disease where major drug targets like cholinesterases (14) and g-secretase (15; 16) are dynamically regulated by their substrates producing complex interactions with potential drug candidates.

Figures

Figure 1. Schematic representing potential catalytically active enzyme states produced by increasing substrate association (S), along with states induced by increasing modifier association (M) which could be either inhibitors or activators.

Figure 2. Schematic representing the modular expansion of the (a) Michaelis-Menten equation through the summation of binding curves. In this example (b) secondary binding results in substrate activation ($V_2 > V_1$), followed by (c) substrate inhibition ($V_1 > V_3$) upon the binding of the third substrate.

Figure 3. Fitting the SHP2 inhibition by AA dataset to a one substrate one modifier model results in four possible fits, (a) a fit where the Hill coefficient is not applied, (b) a fit where the Hill coefficient is applied to the binding of the substrate, (c) a fit where the Hill coefficient is applied to the binding of the inhibitor and (d) fitting of the model with Hill coefficients applied to both binding interactions. The data set represents the substrate range between 17 μM to 20 mM p-nitrophenyl phosphate with 0 μM (blue), 100 μM (orange), 200 μM (green), 300 μM (yellow) and 400 μM (red) AA. Fittings are accompanied by correlation plots of the calculated versus the observed velocities.

Figure 4. Fitting the of the SHP2 inhibition by AA dataset to the best fit models as defined by the minimum AIC value for the model where (a) only the non-substrate activated enzyme interacts with the inhibitor (b) the inhibitor interacts with both substrate hydrolyzing forms of the enzyme (c) the minimum AIC of the examined models and (d) the minimum sum of squared residuals of the examined models. The data set represents the substrate range between 17 μM to 20 mM p-nitrophenyl phosphate with 0 μM (blue), 100 μM (orange), 200 μM (green), 300 μM (yellow) and 400 μM (red) AA. Fittings are accompanied by correlation plots of the calculated versus the observed velocities.

Table 1. Sum of squared residuals and AIC values associated with the models fit to the SHP2 AA

inhibition data.

SSR					ΔAIC					
		S1	S1h		S1	S1h				
a	x1	0.01678	0.01247		22.43	19.27				
	x1h	0.00899	0.00420		13.59	2.38				
b	x1 xx1	0.01581	0.01108		27.39	23.22				
	x1h xx1	0.01579	0.00267		29.37	0.51				
	x1 xx1h	0.00797	0.00268		17.50	0.53				
	x1h xx1h	0.00767	0.00246		18.83	1.08				
		S1 S2	S1h S2	S1 S2h	S1h S2h	S1 S2h	S1h S2h			
c	x1	0.01580	0.01283	0.01435	0.01257	25.39	23.76	25.71	25.42	
	x1h	0.00644	0.00374	0.00515	0.00365	11.79	4.33	11.35	7.15	
d	x2	0.03633	0.01440	0.01359	0.01361	39.85	25.77	24.77	26.79	
	x2h	0.00732	0.00636	0.00395	0.00395	14.01	13.57	5.31	7.30	
e	x1 x2	0.01077	0.01045	0.01077	0.01034	24.72	26.20	26.72	28.02	
	x1h x2	0.00394	0.00308	0.00308	0.00306	9.24	6.96	7.00	8.89	
	x1 x2h	0.00381	0.00380	0.00334	0.00215	8.69	10.65	8.40	11.19	
	x1h x2h	0.00248	0.00227	0.00311	0.00224	3.21	3.65	9.16	5.44	
f	x1 xx1	0.01496	0.00650	0.01465	0.00290	30.44	17.94	32.07	5.94	
	x1h xx1	0.01301	0.00215	0.00662	0.00214	30.01	0.71	20.27	2.63	
	x1 xx1h	0.00427	0.00227	0.00368	0.00226	12.66	1.66	10.08	3.64	
	x1h xx1h	0.00251	0.00184	0.00300	0.00176	3.40	0.00	8.50	1.23	
g	x2 xx2	0.03105	0.00567	0.01240	0.00368	43.12	15.58	29.17	27.30	
	x2h xx2	0.02430	0.00272	0.00556	0.00343	40.86	4.83	17.24	10.87	
	x2 xx2h	0.01413	0.00297	0.01273	0.00225	31.45	6.37	31.64	3.50	
	x2h xx2h	0.01601	0.00229	0.01069	0.00205	35.62	3.80	30.60	3.94	
h	x1 xx1 x2	0.01148	0.00206	0.00689	0.00189	31.83	3.98	24.98	4.47	
	x1h xx1 x2	0.00494	0.00198	0.00396	0.00181	19.19	5.31	17.36	5.71	
	x1 xx1h x2	0.00239	0.00214	0.00201	0.00151	6.60	6.61	5.55	8.61	
	x1 xx1 x2h	0.00476	0.00204	0.00316	0.00188	18.53	5.82	13.43	7.17	
	x1h xx1h x2	0.00203	0.00178	0.00179	0.00177	5.72	5.43	5.57	7.34	
	x1h xx1 x2h	0.00302	0.00198	0.00277	0.00180	12.63	13.48	13.11	14.09	
	x1 xx1h x2h	0.00226	0.00205	0.00175	0.00157	7.59	7.90	5.13	5.27	
	x1h xx1h x2h	0.00176	0.00176	0.00169	0.00148	5.27	6.01	6.59	6.29	
i	x1 x2 xx2	0.01110	0.01110	0.01105	0.00638	31.24	33.12	33.18	25.63	
	x1h x2 xx2	0.00308	0.00306	0.00236	0.00226	10.96	12.86	8.38	10.52	
	x1 x2h xx2	0.00710	0.00710	0.00264	0.00242	25.49	27.22	10.28	10.75	
	x1 x2 xx2h	0.00368	0.00330	0.00228	0.00213	14.09	14.17	7.75	8.61	
	x1h x2h xx2	0.00271	0.00186	0.00184	0.00172	10.79	6.21	6.03	6.84	
	x1h x2 xx2h	0.00237	0.00223	0.00235	0.00194	8.42	9.33	10.30	8.94	
	x1 x2h xx2h	0.00288	0.00257	0.00250	0.00248	11.84	11.83	11.35	13.20	
	x1h x2h xx2h	0.00175	0.00168	0.00175	0.00168	5.20	6.41	7.17	8.41	
j	x1 xx1 x2 xx2	0.01089	0.01073	0.00591	0.00564	36.91	38.67	28.31	29.48	
	x1h xx1 x2 xx2	0.01089	0.01075	0.00591	0.00558	38.91	40.69	30.30	31.30	
	x1 xx1h x2 xx2	0.00229	0.00222	0.00215	0.00179	11.82	13.33	12.75	11.55	
	x1 xx1 x2h xx2	0.00739	0.00601	0.00739	0.00591	32.19	30.60	34.19	32.30	
	x1 xx1 x2 xx2h	0.00307	0.00281	0.00261	0.00243	16.95	17.39	16.12	16.83	
	x1h xx1h x2 xx2	0.00153	0.00124	0.00125	0.00115	6.85	5.17	5.37	5.82	
	x1h xx1 x2h xx2	0.00744	0.00636	0.00730	0.00205	34.29	33.57	35.97	15.87	
	x1h xx1 x2 xx2h	0.00297	0.00280	0.00244	0.00217	18.37	19.30	16.92	16.93	
	x1 xx1h x2h xx2	0.00215	0.00204	0.00197	0.00179	12.74	13.80	13.18	13.52	
	x1 xx1h x2 xx2h	0.00177	0.00134	0.00173	0.00132	9.38	6.58	10.91	8.26	
	x1 xx1 x2h xx2h	0.00293	0.00270	0.00270	0.00239	18.11	18.68	18.66	18.56	
	x1h xx1h x2h xx2	0.00129	0.00126	0.00125	0.00117	5.92	7.41	7.31	8.23	
	x1h xx1h x2 xx2h	0.00135	0.00129	0.00124	0.00114	6.59	7.81	7.23	7.69	
	x1h xx1 x2h xx2h	0.00241	0.00226	0.00220	0.00198	16.70	17.60	17.16	17.29	
	x1 xx1h x2h xx2h	0.00144	0.00121	0.00141	0.00118	7.79	6.79	9.35	8.27	
	x1h xx1h x2h xx2h	0.00125	0.00117	0.00125	0.00113	7.26	8.10	9.25	9.55	

The table is divided based on the inclusion of substrate and modifier induced enzyme states in the model. The columns are defined by substrate-induced forms where S1 indicates one substrate interaction, S2 indicates two and an h indicates the inclusion of Hill coefficients. Rows are defined

by potential inhibitor interactions where the number of x's indicated the number of binding interactions, the number indicates the substrate-induced enzyme form the inhibitor is binding and an h indicates the inclusion of a Hill coefficient. For example, xx2h indicates a second inhibitor binding to the substrate activated enzyme where a Hill modifier is included. The color code is used to highlight better fits (green) from poorer fits (red).

AUTHOR INFORMATION

Corresponding Author

Ryan Walsh PhD

Armand-Frappier Santé Biotechnologie Research Centre, Institut national de la recherche scientifique, 531 boulevard des Prairies, Laval (Quebec), H7V 1B7, Canada.

Ronin Institute, 127 Haddon Pl, Montclair, New Jersey, 07043 USA.
ryan.walsh@ronininstitute.org

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Funding Sources

The Authors did not receive funding for this work.

Notes

ABBREVIATIONS

AA, abietic acid; Δ AIC, change in Akaike information criterion; SHP2, protein tyrosine phosphatase nonreceptor type 11; SSR, sum of squared residuals.

REFERENCES

1. Buker, S. M.; Boriack-Sjodin, P. A.; Copeland, R. A. Enzyme–Inhibitor Interactions and a Simple, Rapid Method for Determining Inhibition Modality. *SLAS DISCOVERY: Advancing Life Sciences R&D* 2019, 24, 515-522.
2. Walsh, R.; Martin, E.; Darvesh, S. A versatile equation to describe reversible enzyme inhibition and activation kinetics: Modeling β -galactosidase and butyrylcholinesterase. *Biochimica et Biophysica Acta (BBA)-General Subjects* 2007, 1770, 733-746.
3. Walsh, R. Comparing enzyme activity modifier equations through the development of global data fitting templates in Excel. *PeerJ* 2018, 6, e6082.
4. Walsh, R. Alternative perspectives of enzyme kinetic modeling. In *Medicinal chemistry and drug design.*; Ekinici, D., Ed.; InTech: London, 2012; pp 357-372.
5. Hjortness, M. K.; Riccardi, L.; Hongdusit, A.; Ruppe, A.; Zhao, M.; Kim, E. Y.; Zwart, P. H.; Sankaran, B.; Arthanari, H.; Sousa, M. C.; De Vivo, M. Abietane-type diterpenoids inhibit protein tyrosine phosphatases by stabilizing an inactive enzyme conformation. *Biochemistry* 2018, 57, 5886-5896.
6. Jia, Z.; Barford, D.; Flint, A. J.; Tonks, N. K. Structural basis for phosphotyrosine peptide recognition by protein tyrosine phosphatase 1B. *Science*, 1995, 268, 1754-1758.
7. Caillet-Saguy, C.; Toto, A.; Guerois, R.; Maisonneuve, P.; di Silvio, E.; Sawyer, K.; Gianni, S.; Wolff, N. Regulation of the Human Phosphatase PTPN4 by the inter-domain linker connecting the PDZ and the phosphatase domains. *Scientific reports*, 2017, 7, 1-9.

8. Srinivasan, B.; Marks, H.; Mitra, S.; Smalley, D. M.; Skolnick, J. Catalytic and substrate promiscuity: distinct multiple chemistries catalysed by the phosphatase domain of receptor protein tyrosine phosphatase. *Biochemical Journal*, 2016, 473, 2165–2177.
9. LaRochelle, J. R.; Fodor, M.; Vemulapalli, V.; Mohseni, M.; Wang, P.; Stams, T.; LaMarche, M. J.; Chopra, R.; Acker, M.G.; Blacklow, S. C. Structural reorganization of SHP2 by oncogenic mutations and implications for oncoprotein resistance to allosteric inhibition. *Nature Communications*, 2018, 9, 4508.
10. Kemmer, G.; Keller, S. Nonlinear least-squares data fitting in Excel spreadsheets. *Nature Protocols* 2010, 5, 267-281.
11. Walsh, R. A reanalysis of Protein Tyrosine Phosphatases Inhibitory studies using the unnatural substrate analogue p-nitrophenyl phosphate. *Anal. Biochem.* 2019, 572, 58-62.
12. Burnham, K. P.; Anderson, D. R. *Model Selection and Multimodel Inference: a Practical Q3 Information-Theoretic Approach*, Springer-Verlag: New York, Inc, 2002.
13. Wang, W.; Liu, L.; Song, X.; Mo, Y.; Komma, C.; Bellamy, H. D.; Zhao, Z. J.; Zhou, G. W. Crystal structure of human protein tyrosine phosphatase SHP-1 in the open conformation. *Journal of cellular biochemistry* 2011, 112, 2062-2071.
14. Walsh, R.; Rockwood, K.; Martin, E.; Darvesh, S. Synergistic inhibition of butyrylcholinesterase by galantamine and citalopram. *Biochim Biophys Acta.* 2011, 1810, 1230-1235.
15. Svedružić, Ž. M.; Popović, K.; Šendula-Jengiđ, V. Modulators of γ -secretase activity can facilitate the toxic side-effects and pathogenesis of Alzheimer's disease. *PloS one* 2013, 8.

16. Walsh, R. Are improper kinetic models hampering drug development? PeerJ 2014, 2, e649.