Published in final edited form as:

J Am Chem Soc. 2018 March 14; 140(10): 3483–3486. doi:10.1021/jacs.7b10368.

Free-Energy-Based Protein Design: Re-Engineering Cellular Retinoic Acid Binding Protein II Assisted by the Moveable-Type Approach

Haizhen A. Zhong^{†,‡}, Elizabeth M. Santos[†], Chrysoula Vasileiou[†], Zheng Zheng[†], James H. Geiger[†], Babak Borhan[†], Kenneth M. Merz Jr.^{*,†}

[†] Department of Chemistry, Michigan State University, East Lansing, Michigan 48824, United States

[‡] Department of Chemistry, University of Nebraska at Omaha, Omaha, Nebraska 68182, United States

Abstract

How to fine-tune the binding free energy of a small-molecule to a receptor site by altering the amino acid residue composition is a key question in protein engineering. Indeed, the ultimate solution to this problem, to chemical accuracy (± 1 kcal/mol), will result in profound and wide-ranging applications in protein design. Numerous tools have been developed to address this question using knowledge-based models to more computationally intensive molecular dynamics simulations-based free energy calculations, but while some success has been achieved there remains room for improvement in terms of overall accuracy and in the speed of the methodology. Here we report a fast, knowledge-based movable-type (MT)-based approach to estimate the absolute and relative free energy of binding as influenced by mutations in a small-molecule binding site in a protein. We retrospectively validate our approach using mutagenesis data for retinoic acid binding to the Cellular Retinoic Acid Binding Protein II (CRABP II) system and then make prospective predictions that are borne out experimentally. The overall performance of our approach is supported by its success in identifying mutants that show high or even sub-nanomolar binding affinities of retinoic acid to the CRABP II system.

The ability to determine which residues to modify in a protein to optimize a target end point (enhanced activity, ligand binding affinity, protein stability, etc.) is key to further advancing protein engineering. The routine and accurate answer to this question has applications in protein design to improve function,^{1,2} in the design of protein probes with a range of applications (e.g., fluorescent protein tag³), etc. Many knowledge-based and physics-based force field approaches have been developed to assist in identifying appropriate residues for mutational studies.^{1–5}

*Corresponding Author kmerz1@gmail.com.

The authors declare no competing financial interest.

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.7b10368.

Computational procedures in setting up proteins and ligands for free energy calculations, and experimental methods for expression, purification and characterization of binding constants (PDF)

The main challenges for this field are the concomitant need to sample relevant conformational space sufficiently while also computing the energies of these states to good accuracy. Even with these challenges there have been many successes.^{1–5} Herein we report the use of a fast and accurate approach to estimate changes in the binding free energies of mutant proteins to a small-molecule.

In this work, we have applied the newly developed free energy method, the “movable-type” (MT) method, to perform end-state free energy simulations of the CRABP II protein system. The unique attribute of the MT method is that it uses numerical approximations to extrapolate the local partition function centered on an initial or “seed” structure. The MT method uses the approximation that all atom pairwise potentials with respect to each atom are independent in the close neighborhood of a given conformation. The full numerical details of this approach are given in the extant literature.^{6–8} This method is relatively fast, only taking several minutes to obtain estimates for the free energy of binding for each protein–ligand complex. For example, this method has been successful in estimating the experimental binding constants for large collections of protein/ligand complexes⁷ and solvation free energies.⁸

In this paper, we apply this approach to explore the effect of mutating residues on the binding affinity of retinoic acid (RA) to CRABP II (Figure 1). The advantage of this approach is that it is readily applicable to engineer proteins containing several mutations at the same time, moving beyond, for example, the alanine scanning method.^{9,10} We first show retrospectively how it performs and then follow this by a prospective challenge to predict novel mutations for subsequent validation experiments.

CRABP II is a small cytosolic protein that binds all-transretinoic acid. Borhan and co-workers have re-engineered CRABP II to generate rhodopsin mimics^{11,12} and a colorimetric pH sensor¹³ via covalent binding of all-trans-retinal via protonated Schiff base formation. The binding of retinoic acid (RA) to CRABP II is slightly different from that of retinal (RT) binding in that retinal forms a covalent bond with Lysine in the R132 K:R111L:L121E triple mutant. The covalent bond makes the terminal carbon (imine carbon in RT, versus carboxylate carbon in RA) shift 1.9 Å from the carboxylate carbon in RA.¹⁴

To test the applicability of MT in guiding protein design, Borhan’s group provided a list of CRABP II mutants along with experimentally determined binding constants for RA binding to CRABP II mutants (Table 1). We used crystal structure 2FR3 (wild-type) as the template to build the double mutants,¹⁵ 2G7B (triple mutant R111L:L121E:R132K) to build the triple and quadruple mutants,¹² and 3CWK (penta mutant R132K:Y134F:R111L:T54V:L121E)¹⁶ to build the penta mutants listed in Table 1. The reason for selecting these PDB files as templates is to minimize the potential structural effect on ligand binding due to multiple mutations. To evaluate the dependency of initial PDB structures on G_{bind} estimation, we also built all mutants using the WT structure 2FR3 and the estimated G_{bind} for these mutants are reported in the SI. At the same time, we also included in this study the crystal structures of the CRABP II mutants or wild-type proteins (2CBS, 3CBS, 4I9R, 4I9S, and 2G79) with bound RA analogs (R12, R13 or retinal RT) (see Table 1).

All X-ray crystal structures were downloaded from the Protein Databank and were prepared with established protein preparation procedures (see SI for full details), followed by minimizations using the MacroModel module within the Schrödinger software suite where the side chains and ligands were relaxed to reduce steric clash. The minimized protein–ligand complexes were then saved for input into the MT-program (a MatLab-based program). The output of the MT-program is given as a pK_d and as a free energy of binding G .

Table 1 shows that the MT performed well in predicting the binding constants of mutant proteins for 39 CRABP II protein systems: the Pearson's correlation coefficient was 0.73 and the R^2 was 0.53 (Table 2). The root-mean-square (RMS) error and the mean of absolute error (MAE) for the absolute binding free energy were around 2.7 kcal/mol lower than the observed ones. However, the relative binding free energies (using the R132 K:E73A double mutant as reference for all mutants) were very small (around 0.6–0.7 kcal/mol) indicating that the free energy spacing among the mutants was well reproduced. The dependency of the initial PDB structures on the G_{bind} was small as the mean error between the predicted G_{bind} from mutants and those from WT was between 0.73 and 0.82 kcal/mol (Tables 5S and 6S).

We next applied MT to estimate ten additional CRABP II mutants with unknown binding constants. We prepared the ten CRABP II mutants (ranging from double to quadruple mutants, see Table 3) by mutating residues as needed and then the free energies were calculated. The choice of these mutants were made based on the available binding constants of retinoic acid binding generated by the experimental group and withheld from the computational team. From the MT-generated G values, predicted experimental G values, also called G_{FIT} values were obtained by using the equation $G_{\text{exp}} = 0.97 G_{\text{pred}} - 2.79$ (Figure 2A). Once we obtained the fitted G values for the mutants listed in Table 3, we communicated our predictions to Borhan's group at which time they provided us with the experimental binding affinities.

Table 3 clearly shows that the predicted G values were able to successfully estimate the experimental data (G_{exp}) with RMS errors and MAE less than 0.7 kcal/mol (G_T in Table 2). In the MT-approach, the majority of time was spent on protein structural preparation and minimization rather than actual simulation, which is inverted relative to, for example, MD-based approaches.¹⁷ In the present study, the calculation of G_{bind} for a given protein/ligand complex took less than a minute on a modern laptop.

Next we undertook a prospective study to improve the binding affinity of CRABP II mutants toward retinoic acid (RA). The binding affinity of CRABP II and all-trans-RA is very high (2.0 ± 1.2 nM), whereas the binding of CRABP II to retinal was 3 0 0 0 - f o l d w e a k e r ($K_d = 6 6 0 0$ nM).¹¹ The R132K:R111L:L121E triple mutant of CRABP II significantly enhances the binding of retinal ($K_d = 1.4$ nM).¹⁴ The replacement of Arg132 with lysine removed the water molecule, thus allowing suitable nucleophilic attack on the carbonyl of retinal to form a Schiff base.¹⁴ Thus, appropriate mutations enhance ligand binding (be it RA or retinal).

In our work, we turned our attention to Val41, Ile52, and Leu121, three hydrophobic residues in the pocket where the RA carboxylate group binds (Figure 1). Table 1 shows that triple mutant R132K:R111L:L121Q exhibited tighter RA binding affinity ($K_d = 306$ nM) than that of the double mutant R132K:R111L ($K_d = 736$ nM). However, the extent of improvement appeared to be associated with the polarity of residue 121 where Gln121 (Q121) improved the binding whereas a negatively charged Asp121 (triple mutant R132K:R111L:L121D, $K_d = 2639$ nM) lowered the binding affinity.

To enhance RA's binding affinity, we proposed that substitution of Val41 with a Gln (V41Q, Figure 3A) would provide H-bonds with Arg111, a residue important for RA binding. Similarly, replacing Ile52 with a Glu (I52E, Figure 3B) formed a H-bond with Arg111. Mutating L121 to Asn (L121N, Figure 3C) or to Gln (L121Q, Figure 3D) would not only provide hydrogen bonds to Arg132 but also form a direct H-bond with the RA carboxylate group.

We systematically mutated Val41, Ile52, and Leu121 and predicted the binding free energies (Figure 1S, SI). Among these 57 mutations, several were promising, but we decided that the single mutants I52E, L121N, L121Q, V41E, and V41Q would be made. The binding affinity, as measured by K_d values showed that all five mutants gave very good binding affinities, two of which bound tighter than WT (2 nM) (Table 4). A further 10 mutants were explored, with 4 not expressing as soluble proteins, 3 showing good agreement between experiment and theory, and 3 where MT predicted an order of magnitude (in K_d) better binding affinity than found experimentally (see SI).

In summary, we have validated, both retrospectively and prospectively, a MT-based free energy of binding approach, in a protein engineering exercise of CRABP II mutants. The efficient MT approach was shown to have good predictive ability and is readily applicable to other protein design projects.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

K.M.M. thanks the NIH (GM112406) for supporting the present research. B.B. thanks the NIH (GM101353) for support. H.A.Z. thanks the Faculty Development Fellowship from University of Nebraska at Omaha.

REFERENCES

- (1). Perez A; Morrone JA; Simmerling C; Dill KA Curr. Opin. Struct. Biol. 2016, 36, 25–31. [PubMed: 26773233]
- (2). London N; Ambroggio XJ Struct. Biol. 2014, 185, 136–146.
- (3). Sandhya S; Mudgal R; Kumar G; Sowdhamini R; Srinivasan N Curr. Opin. Struct. Biol. 2016, 37, 71–80. [PubMed: 26773478]
- (4). Khare SD; Fleishman SJ FEBS Lett. 2013, 587, 1147–1154. [PubMed: 23262222]
- (5). Yang W; Lai L Curr. Opin. Struct. Biol. 2017, 45, 67–73. [PubMed: 27951448]
- (6). Zheng Z; Merz KM, J. Jr Chem. Inf. Model. 2013, 53, 1073–1083.
- (7). Zheng Z; Ucisik MN; Merz KM, J. Jr Chem. Theory Comput. 2013, 9, 5526–5538.

- (8). Zheng Z; Wang T; Li P; Merz KM, J. Jr Chem. Theory Comput. 2015, 11, 667–682.
- (9). Weiss GA; Watanabe CK; Zhong A; Goddard A; Sidhu SS Proc. Natl. Acad. Sci. U. S. A. 2000, 97, 8950–8954. [PubMed: 10908667]
- (10). Morrison KL; Weiss GA Curr. Opin. Chem. Biol. 2001, 5, 302–307. [PubMed: 11479122]
- (11). Crist RM; Vasileiou C; Rabago-Smith M; Geiger JH; Borhan BJ Am. Chem. Soc. 2006, 128, 4522–4523.
- (12). Vasileiou C; Vaezslami S; Crist RM; Rabago-Smith M; Geiger JH; Borhan BJ Am. Chem. Soc. 2007, 129, 6140–6148.
- (13). Berbasova T; Nosrati M; Vasileiou C; Wang W; Lee KS; Yapici I; Geiger JH; Borhan BJ Am. Chem. Soc. 2013, 135, 16111–16119.
- (14). Vasileiou C; Wang W; Jia X; Lee KS; Watson CT; Geiger JH; Borhan B Proteins: Struct., Funct., Genet 2009, 77, 812–822. [PubMed: 19603486]
- (15). Vaezslami S; Mathes E; Vasileiou C; Borhan B; Geiger JH J. Mol. Biol. 2006, 363, 687–701. [PubMed: 16979656]
- (16). Vaezslami S; Jia X; Vasileiou C; Borhan B; Geiger JH Acta Crystallogr., Sect. D: Biol. Crystallogr. 2008, 64, 1228–1239. [PubMed: 19018099]
- (17). Buch I; Giorgino T; De Fabritiis G Proc. Natl. Acad. Sci. U. S. A. 2011, 108, 10184–10189. [PubMed: 21646537]

Figure 1.
Retinoic acid (RA) bound to wild-type CRABP1. Residues forming H-bonds with the RA are Arg132 and Tyr134.

Figure 2.

Plots of MT-calculated ΔG for the model systems in Table 1 (top, A), and for the systems in Table 1 with an additional five proposed mutants that were subsequently engineered (bottom, B).

Figure 3.
Proposed mutants (cyan) in the CRABP II binding pocket with RA as ligand.

Table 1. Predicted Free Energies (kcal/mol) of Binding for Ligands (retinoic acid unless specified) Bound to CRABP II Mutants^a

Proteins	K_d (nM)	G_{WT} (kcal/mol)	G_{exp} (kcal/mol)
WT-R13 (2CBS)	6	-8.00	-11.22
WT-R12 (3CBS)	58	-6.89	-9.87
R111K:R132L:Y134F:T54V:R59W:A32W-RT (419R)	112	-7.12	-9.48
R111K:R132L:Y134F:T54V:R59W-RT (419S)	162	-6.30	-9.26
R132K:Y134F-RT (2G79)	120	-6.39	-9.44
(2FR3)			
R132K:E73A	564	-5.54	-8.52
R132K:Y134F:T54V	565	-5.32	-8.52
R132K:Y134F:L121E	1400	-5.96	-7.99
R132K:I52D	2742	-5.44	-7.59
R132K:W109L	3196	-5.69	-7.50
R132K:E73A:L121E	353	-5.90	-8.80
(2G7B)			
Y134F:R111L:L121E	220	-5.32	-9.08
R132K:R111L:L121Q	306	-5.58	-8.89
R132K:R111 K:L121E	486	-5.67	-8.61
R132K:Y134F:R111E	530	-5.40	-8.56
R132K:R111E:L121E	608	-5.57	-8.48
R132K:R111M	699	-5.37	-8.40
R132K:R111L	736	-5.39	-8.37
R132K:R111M:L121E	742	-5.65	-8.36
R132K:Y134F:R111L	1000	-5.41	-8.18
R132K:R111E	1088	-5.41	-8.13
R132K:R111H	1362	-5.17	-8.00
R132K:R111V:L121E	2313	-5.69	-7.69
R132K:R111L:L121D	2639	-5.87	-7.61
R132K:R111L:C130D	6105	-5.60	-7.11
R132K:R111L:Y134D	7419	-5.48	-7.00

Proteins	K_d (nM)	G_{NT} (kcal/mol)	G_{exp} (kcal/mol)
(3CWK)			
R132K:Y134F:R111L:L121E:T54V	250	-6.54	-9.01
R132K:Y134F:R111L:L121Q	240	-6.33	-9.03
R132K:R111L:L121E:T54 V	400	-6.67	-8.73
R132K:Y134F:R111L:L121N:T54V	420	-6.08	-8.70
R132K:R111L:L121E:V41E	490	-5.31	-8.61
R132K:Y134F:R111L:L121D:T54V	760	-6.09	-8.35
R132K:Y134F:R111L:T54 V	900	-5.13	-8.25
R132K:Y134F:R111L:L121Q:T54V	1739	-5.34	-7.86
R132K:Y134F:R111L:T54E	2180	-6.29	-7.72
R132K:R111L:L121E:Y134D	2718	-5.19	-7.59
R132K:Y134F:R111L:L121N	3050	-5.28	-7.52
R132K:R111L:L121E:Y134E	3297	-5.20	-7.48
R132K:R111L:L121E:Y134E:T54V	3665	-5.33	-7.42

^a Mutant models were modified based on the protein in parentheses above.

Table 2.Statistical Results for the MT-Based Estimation of Free Energy of Binding for CRABPII Systems^a

	G_A	G_A	G_B	G_B	G_T
No. of complexes	39	39	44	44	10
RMSE (kcal/mol)	2.65	0.68	2.76	0.66	0.65
MAE (kcal/mol)	2.59	0.51	2.69	0.51	0.58
Pearson's R	0.73	0.73	0.81	0.81	0.64
Correlation R^2	0.53	0.53	0.66	0.66	0.41

^aThe G_A values are the free energies for the 39 complex test set (set A) whereas the G_B values are for an additional 5 complexes added to the original 39 (set B). G_T is the predicted free energies of binding for the unknown set.

Table 3.

Prediction of Free Energies of Binding of Retinoic Acid to Tested CRABP II Mutants

Proteins	G_{MT1}	G_{FIT}	G_{exp}	K_d (nM)
R132K:R111K	-5.33	-7.94	-7.23	5016
R132K:L121E	-5.87	-8.46	-8.05	1260
R132K:R111H:L121E	-5.74	-8.33	-8.97	264
R132K:W109L:L121E	-6.06	-8.64	-9.59	93
R132K:Y134F:R111L:L121E	-5.35	-7.96	-8.34	770
R132K:Y134F:R111L:L121D	-5.99	-8.58	-9.14	200
R132K:Y134F:R111E:T54V	-5.20	-7.81	-8.27	860
R132K:R111L:L121Q:T54V	-5.28	-7.89	-7.96	1458
R132L:Y134F:R111L:L121E	-6.21	-8.78	-9.27	160
R132 K:R111L:L121E:C130D	-5.24	-7.85	-8.99	258

Table 4.Prediction and Experimental K_d Values to RA Binding to Five Mutants of CRABP II

Mutants	G_{MT1}	G_{FIT}	K_{dPred}	K_{dExp} (nM)
I52E	-6.95	-9.53	97.41	6.01
L121N	-6.88	-9.47	108.68	0.12
L121Q	-7.11	-9.69	74.38	4.70
V41E	-6.76	-9.35	132.23	7.79
V41Q	-6.80	-9.39	123.85	1.19
WT	-6.74	-9.33	137.42	2.00