Towards Rational Design of Selective Molecularly Imprinted Polymers (MIPs) for Proteins: Computational and Experimental Studies of Acrylamide-Based Polymers for Myoglobin

Mark V. Sullivan, Sarah R. Dennison, Georgios Archontis, Subrayal M. Reddy,*

Joseph M. Hayes*

* Corresponding authors
tel: +441772893291 (SMR); fax: +441772894981; email: smreddy@uclan.ac.uk
tel: +441772894334 (JMH); fax: +441772894981; email: jhayes@uclan.ac.uk

Table of Contents

1. Statistical Thermodynamics

In the SI we present a detailed derivation of the expressions used to compute the monomer binding probabilities and the average number of bound monomers. The related formulas are usually expressed in terms of binding constants. In contrast, here, we employ estimates of the binding free energies ΔG_{PL}.

1.1 Protein with M binding sites.

We consider a protein P with M independent, non-equivalent sites, indexed by the integers $i = \{1,...,M)$. We assume that each protein site can be empty or bind one monomer L molecule. Binding of monomer L at site i is associated with a change in the standard-state free energy change $\Delta G_{PL}(i)$. The resulting grand-canonical partition function system is (*1*)

$$\Xi_{PL} = \prod_{j=1}^{M} \left(1 + x e^{-\beta \Delta G_{PL}(j)}\right) \qquad \text{(SI-1)}$$

where $x \equiv [L]/C_0$, with $[L]$ the *free* monomer concentration in solution and $C_0 = 1 \frac{\text{mol}}{\text{L}} = \frac{1}{1660 \text{ A}^3}$ the standard-state concentration. The quantity ΔG_{PL} is defined in Eq. 1 of the main text.

The average number of monomer L molecules bound to a protein molecule is

$$\overline{M}_L(x) = \sum_{i=1}^{M} \frac{x e^{-\beta \Delta G_{PL}(i)}}{1 + x e^{-\beta \Delta G_{PL}(i)}} \qquad \text{(SI-2)}$$

To employ Eq. (SI-2), we need to know the concentration x of *free* monomer. Since each protein molecule binds on average $\overline{M}_L(x)$ ligand molecules, the concentration of

bound monomer is $[L]_{bound} = [P]_{tot}\,\overline{M}_L(x)$, where $[P]_{tot}$ is the total protein concentration. The concentration of free monomer is then:

$$[L] = [L]_{tot} - [L]_{bound} = [L]_{tot} - [P]_{tot}\,\overline{M}_{PL}(x)$$

with $[L]_{tot}$ the total monomer concentration in solution. Using Eq. (SI-2), we obtain:

$$\frac{[L]}{C_0} = x = \frac{[L]_{tot}}{C_0} - \frac{[P]_{tot}}{C_0}\sum_{i=1}^{M}\frac{x e^{-\beta\Delta G_{PL}(i)}}{1 + x e^{-\beta\Delta G_{PL}(i)}} \qquad \text{(SI-3)}$$

A similar equation has been derived for equivalent sites in reference (*2*). Eq (SI-3) can be solved self-consistently to yield the concentration $x = x^*$ of free monomer in solution. The solution $x = x^*$ can then be used in eq. (SI-2), to compute the average number of bound monomers at a protein molecule.

1.2 Solutions with two types of monomers

Suppose that the solution contains two different monomers L_1 and L_2 . These monomers can bind at the same protein sites with free energy changes $\{\Delta G_{PL_1}(i)\}$ and $\{\Delta G_{PL_2}(i)\}$, respectively. The grand-canonical partition function is now

$$\Xi_{PL_1L_2} = \prod_{i=1}^{M}\left(1 + x_1 e^{-\beta\Delta G_{PL_1}(i)} + x_2 e^{-\beta\Delta G_{PL_2}(i)}\right) \qquad \text{(SI-4)}$$

[compare with expression (SI-1)].

The average numbers of monomers of types L_1 or L_2, bound to a protein molecule, are

$$\overline{M}_{L_1}(x_1) = \sum_{i=1}^{M}\frac{x_1 e^{-\beta\Delta G_{PL_1}(i)}}{1 + x_1 e^{-\beta\Delta G_{PL_1}(i)} + x_2 e^{-\beta\Delta G_{PL_2}(i)}} \qquad \text{(SI-5a)}$$

and

$$\overline{M}_{L_2}(x_2) = \sum_{i=1}^{M} \frac{x_2 e^{-\beta\Delta G_{PL_2}(i)}}{1 + x_1 e^{-\beta\Delta G_{PL_1}(i)} + x_2 e^{-\beta\Delta G_{PL_2}(i)}} \qquad \text{(SI-5b)}$$

By analogy with Eq. (SI-3), the free-ligand concentrations can be evaluated by solving self-consistently the system of equations:

$$\begin{aligned} x_1 &= \frac{[L_1]_{tot}}{C_0} - \frac{[P]_{tot}}{C_0} \sum_{i=1}^{M} \frac{x_1 e^{-\beta\Delta G_{PL_1}(i)}}{1 + x_1 e^{-\beta\Delta G_{PL_1}(i)} + x_2 e^{-\beta\Delta G_{PL_2}(i)}} \\ x_2 &= \frac{[L_2]_{tot}}{C_0} - \frac{[P]_{tot}}{C_0} \sum_{i=1}^{M} \frac{x_2 e^{-\beta\Delta G_{PL_2}(i)}}{1 + x_1 e^{-\beta\Delta G_{PL_1}(i)} + x_2 e^{-\beta\Delta G_{PL_2}(i)}} \end{aligned} \qquad \text{(SI-6)}$$

The solutions $x_1 = x_1^*$ and $x_2 = x_2^*$ can be substituted in Eqs. (SI-5), to yield the average monomers of types L_1 and L_2 bound to a protein molecule. Note that in general the free monomer concentrations differ ($x_1^* \neq x_2^*$), even if the solutions are prepared with equal total monomer concentrations $[L_1]_{tot} = [L_2]_{tot}$.

The probability for site i to be occupied by monomer L_1, regardless of the occupancy state of other sites, is given by:

$$p(\text{ site i is occupied by } L_1\ ) = \frac{x_1 e^{-\beta\Delta G_{PL_1}(i)}}{1 + x_1 e^{-\beta\Delta G_{PL_1}(i)} + x_2 e^{-\beta\Delta G_{PL_2}(i)}} \qquad \text{(SI-7a)}$$

Similarly, the probability for site i to be occupied by monomer L_2, regardless of the occupancy state of other sites, is

$$p(\text{site i occupied by } L_2) = \frac{x_2 e^{-\beta\Delta G_{PL_2}(i)}}{1 + x_1 e^{-\beta\Delta G_{PL_1}(i)} + x_2 e^{-\beta\Delta G_{PL_2}(i)}} \qquad \text{(SI-8b)}$$

The ratio of the above probabilities is

$$\frac{p(\text{site i is occupied by } L_1)}{p(\text{site i is occupied by } L_2)} = \frac{x_1 e^{-\beta\Delta G_{PL_1}(i)}}{x_2 e^{-\beta\Delta G_{PL_2}(i)}} \cong \frac{e^{-\beta\Delta G_{PL_1}(i)}}{e^{-\beta\Delta G_{PL_2}(i)}} \qquad \text{(SI-9)}$$

The last equality holds if the *free* concentrations of the two monomers are approximately equal. Eq. (SI-9) was employed in the calculation of site-specific relative binding probabilities of the two co-monomer solutions reported in Table 2.

2. Tables S1 to S4

Table S1: Glide protein grid coordinate centers in Å, defined by inner box 10 Å × 10 Å × 10 Å, for each binding site.

	X	**Y**	**Z**
Site 1	17.620	-23.778	12.972
Site 2	20.229	-23.142	32.053
Site 3	7.083	-30.952	23.353
Site 4	8.547	-21.480	-2.489
Site 5	2.500	-13.810	7.800
Site 6	2.968	-7.392	27.891
Site 7	24.100	-36.302	20.533
Site 8	26.717	-35.043	11.150
Site 9	5.538	-18.823	34.450
Site 10	-2.100	-29.760	1.650
Site 11	-4.436	-29.653	13.807
Site 12	-5.550	-15.110	18.950
Site 13	4.850	-17.796	18.379
Site 14	14.183	-22.010	27.283

Table S2: The top-ranked GlideScore value for each of the five monomers (AAm, NHMAm, NHEAm, DMAm, TrisNHMAm), docked using Glide-SP into each of the 14 binding sites of myoglobin as predicted by SiteMap.[a]

	Binding Site													
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**
AAm	-1.79	-2.70	-2.17	-2.23	-2.72	-2.51	-2.04	-1.95	-2.14	-2.28	-2.17	-2.00	-1.85	-1.42
NHMAm	-1.01	-1.33	-1.46	-1.10	-0.93	-1.60	-1.41	-0.61	-1.36	-1.46	-2.06	-1.69	**0.37**	*n/a*
NHEAm	*n/a*	-1.54	-1.55	-1.07	-0.93	-1.21	-1.31	-0.78	-1.12	-1.70	-0.49	-1.60	*n/a*	*n/a*
DMAm	*n/a*	-2.74	-2.90	-2.53	-2.17	-2.65	-2.02	-2.03	-1.89	-2.68	-2.08	-1.88	*n/a*	*n/a*
TrisNHAm	-0.31	**0.51**	**0.65**	**0.46**	**0.13**	**0.02**	**0.81**	**0.86**	-0.98	-0.26	1.32	-0.77	*n/a*	*n/a*

[a] GlideScores that are positive are highlighted in bold. *n/a* indicates that no binding poses were obtained.

Table S3: The breakdown of ΔG_{PL} into its contributions for each of the five monomers (AAm, NHMAm, NHEAm, DMAm, TrisNHMAm) calculated using MM-GBSA at each the 14 predicted binding sites of myoglobin.[a]

Monomer	**Binding Site**													
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**
AAm														
ΔE_{PL}^{MM}	-21.76	-32.66	-20.4	-25.89	-33.01	-25.96	-24.70	-18.62	-21.67	-28.98	-15.78	-16.86	-17.46	-20.49
ΔG_{PL}^{solv}	10.65	12.96	5.71	13.01	14.80	12.05	13.63	5.22	9.66	15.17	5.61	5.00	10.65	14.83
$-T\Delta S_{PL}^{RRHO}$	13.59	10.72	10.94	10.65	10.92	11.07	12.65	10.13	10.82	10.21	10.25	10.19	14.12	14.85
ΔG_{PL}	**2.5**	-9.0	-3.8	-2.2	-7.3	-2.8	**1.6**	-3.3	-1.2	-3.6	**0.1**	-1.7	**7.3**	**9.2**
NHMAm														
ΔE_{PL}^{MM}	-21.94	-34.33	-29.5	-13.49	-43.54	-35.21	-31.93	-27.81	-27.04	-33.86	-19.65	-24.13	10.50	*n/a*
ΔG_{PL}^{solv}	6.83	10.84	10.84	16.74	21.45	18.32	18.57	10.71	9.73	11.80	6.69	11.91	9.77	*n/a*
$-T\Delta S_{PL}^{RRHO}$	11.74	12.66	13.56	12.98	12.26	13.98	14.01	11.77	12.47	12.13	12.27	12.74	16.50	*n/a*
ΔG_{PL}	-3.4	-10.8	-5.1	**16.2**	-9.8	-2.9	**0.7**	-5.3	-4.8	-9.9	-0.7	**0.5**	**36.8**	*n/a*
NHEAm														
ΔE_{PL}^{MM}	n/a	-34.85	-34.54	-31.24	-41.55	-36.82	-33.29	-27.53	-27.88	-36.83	-15.97	-29.8	*n/a*	*n/a*
ΔG_{PL}^{solv}	n/a	11.33	13.65	12.96	20.04	18.75	18.76	8.94	11.35	11.52	10.54	14.30	*n/a*	*n/a*
$-T\Delta S_{PL}^{RRHO}$	*n/a*	12.23	13.89	12.38	12.91	14.91	14.01	11.97	12.17	12.42	12.29	12.35	*n/a*	*n/a*
ΔG_{PL}	*n/a*	-11.3	-7.0	-5.9	-8.6	-3.2	-0.5	-6.6	-4.4	-12.9	**6.9**	-3.2	*n/a*	*n/a*
DMAm														
ΔE_{PL}^{MM}	*n/a*	*-34.54*	*-25.67*	*-30.38*	*-29.2*	*-27.69*	*-24.79*	*-23.49*	*-32.66*	*-35.18*	*-29.67*	*-25.79*	*n/a*	*n/a*
ΔG_{PL}^{solv}	*n/a*	14.32	5.18	11.84	10.31	11.84	10.46	6.97	17.32	14.12	15.30	12.78	*n/a*	*n/a*
$-T\Delta S_{PL}^{RRHO}$	*n/a*	11.46	11.35	12.04	11.59	12.01	15.02	11.08	11.08	10.41	11.08	10.70	*n/a*	*n/a*
ΔG_{PL}	*n/a*	-8.8	-9.1	-6.5	-7.3	-3.8	**0.7**	-5.4	-4.3	-10.7	-3.3	-2.3	*n/a*	*n/a*
TrisNHAm														
ΔE_{PL}^{MM}	-38.13	-32.84	-38.49	-42.96	-43.49	-44.76	-25.38	-42.58	-22.26	-45.93	-32.97	-28.51	*n/a*	*n/a*
ΔG_{PL}^{solv}	14.44	12.48	14.44	18.02	17.35	21.63	10.15	18.39	7.20	8.14	14.49	10.33	*n/a*	*n/a*
$-T\Delta S_{PL}^{RRHO}$	13.40	15.17	13.36	14.22	14.03	15.76	14.74	12.90	13.75	12.93	13.32	12.99	*n/a*	*n/a*
ΔG_{PL}	-10.3	-5.2	-10.7	-10.7	-12.1	-7.4	-0.5	-11.3	-1.3	-24.9	-5.2	-5.2	*n/a*	*n/a*

[a] Values that are positive (unfavourable binding) are highlighted in bold. *n/a* indicates that there were no predicted binding poses.

Table S4: The ΔE_{Hbond} contributions for each of the five monomers (AAm, NHMAm, NHEAm, DMAm, TrisNHMAm) calculated using MM-GBSA at each the 14 predicted binding sites of myoglobin.[a]

	Monomer				
Binding Site	**AAm**	**NHMAm**	**NHEAm**	**DMAm**	**TrisNHMAm**
1	-1.02	-0.51	*n/a*	*n/a*	-1.70
2	-1.51	-2.13	-1.97	-0.89	-1.00
3	-1.18	-1.01	-1.02	-0.49	-1.70
4	-1.18	-1.30	-1.34	-0.74	-1.06
5	-1.96	-1.42	-1.34	-0.74	-1.06
6	-1.28	-1.77	-1.66	-1.03	-1.26
7	-1.5	-2.04	-2.25	-1.21	-0.48
8	-0.76	-1.08	-1.09	-0.54	-0.96
9	-2.39	-1.53	-1.66	-2.45	-0.30
10	-1.80	-1.74	-1.71	-1.08	-2.35
11	-1.63	-1.32	-0.79	-0.58	-0.74
12	-1.84	-1.25	-0.99	-1.04	-0.82
13	-2.04	*n/a*	*n/a*	*n/a*	*n/a*
14	-1.83	*n/a*	*n/a*	*n/a*	*n/a*

[a] *n/a* indicates that there were no predicted binding poses.

3. Figure S1.

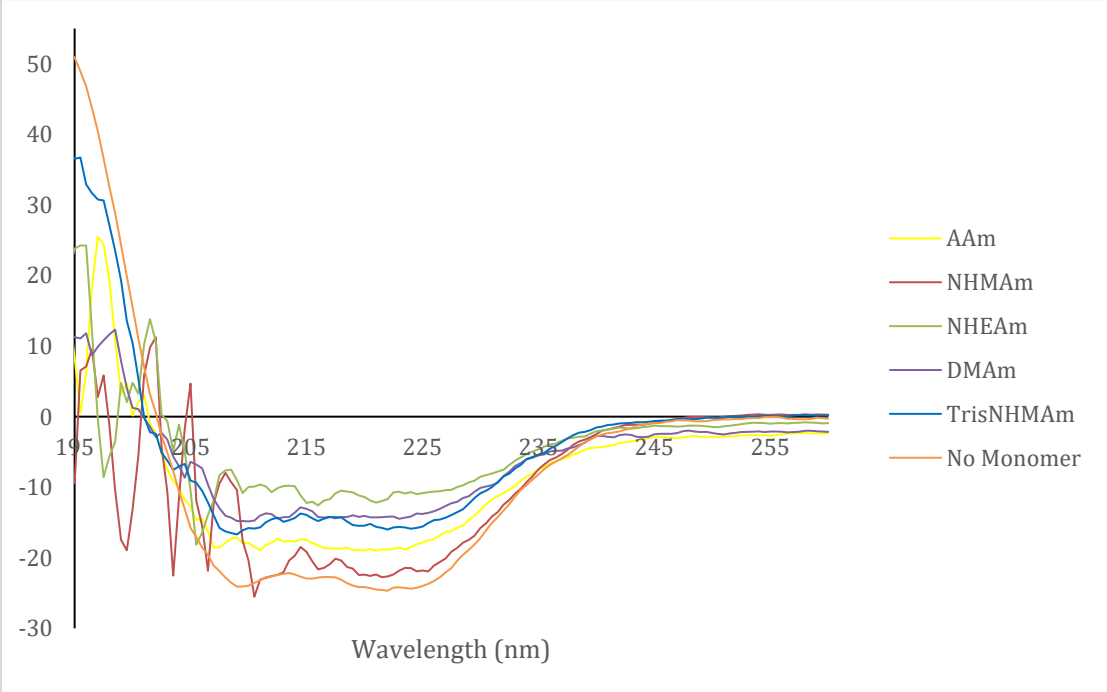

Figure S1: CD spectroscopic analysis of myoglobin after being mixed with the five monomers: AAm (yellow), NHMAm (red), NHEAm (green), DMAm (purple), TrisNHMAm (blue) and no monomer (orange), at a protein:monomer ratio of 1:1081, the same as the polymerisation solution used in hydrogel MIP formation.

4. SI Bibliography

1. Hill, T. Cooperativity Theory in Biochemistry. *Springer Series in Molecular Biology*, **1985**, Series Ed. Alexander Rich.
2. Curk, T.; Dobnikar, J.; Frenkel, D. Rational Design of Molecularly Imprinted Polymers. *Soft Matter* **2016**, *12*, 35-44.