Supporting Information for

Dependence of Avidity on Linker Length for a Bivalent Ligand-Bivalent Receptor Model System

Eric T. Mack, Phillip W. Snyder, Raquel Perez-Castillejos, Başar Bilgiçer, Demetri T. Moustakas, Manish J. Butte, and George M. Whitesides*

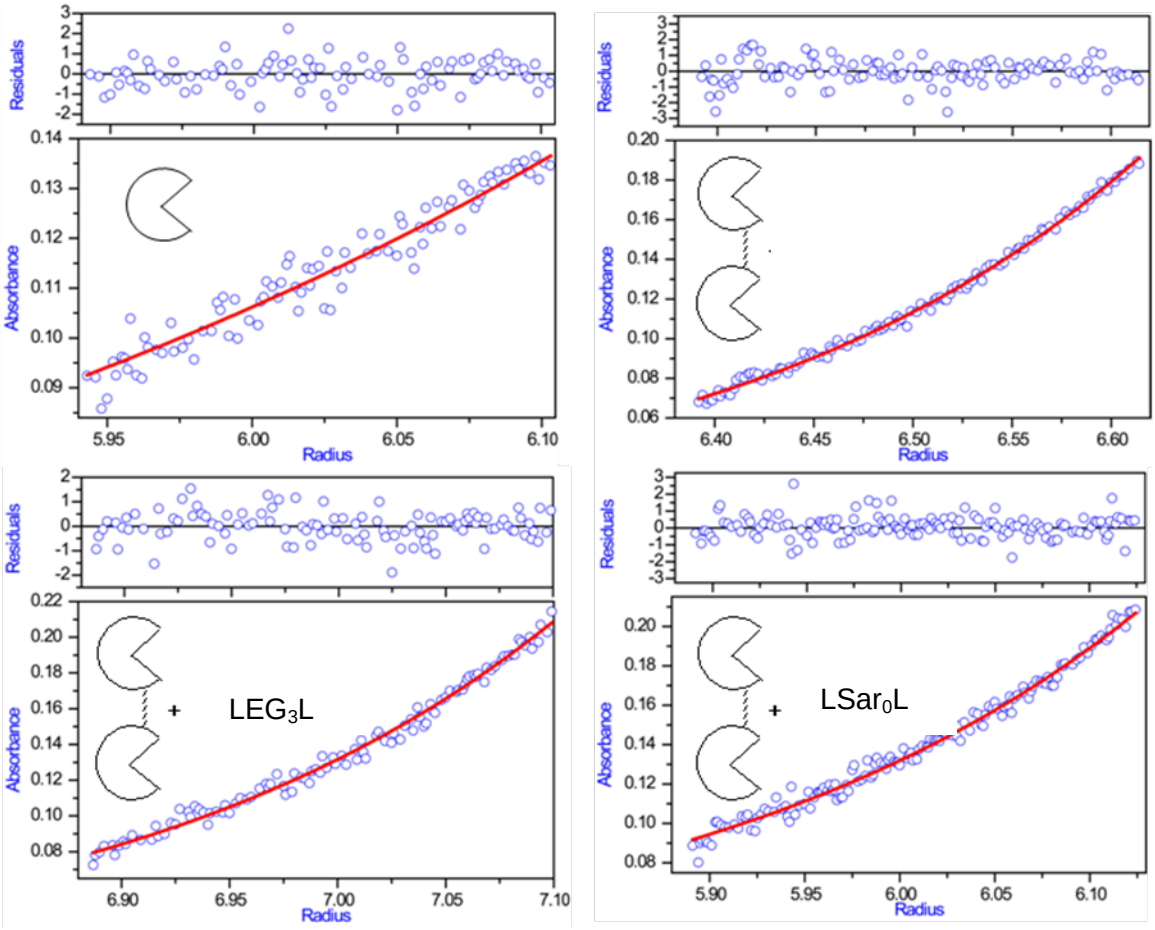
Department of Chemistry and Chemical Biology, Harvard University

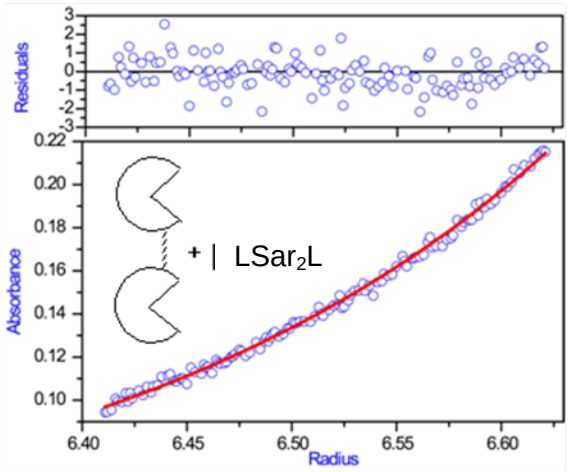
12 Oxford Street, Cambridge, MA 02138

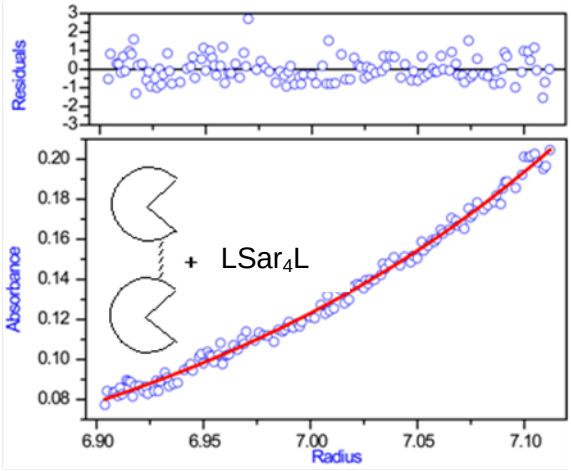
* Correspondence: gwhitesides@gmwgroup.harvard.edu

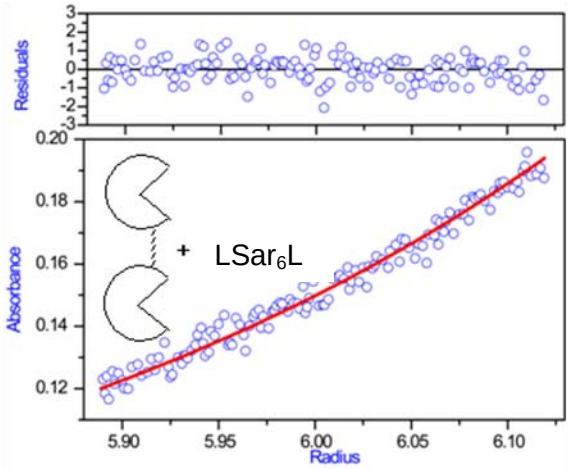
Table of Contents

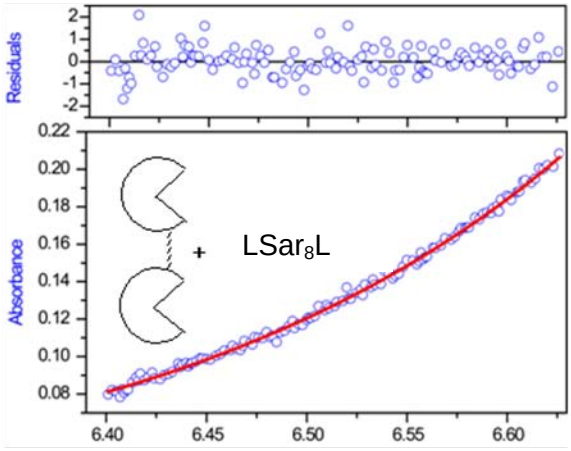
Figure S1. Sedimentation equilibrium experiments of CA and $(CA)_2$ with and without added LRL as observed at 280 nm at 25 ºC. The hollow circles are experimental data for and the line is the fit for a single ideal species.




Residuals
Absorbance
Radius
+ | LSar2L


Residuals
Absorbance
Radius
+ LSar4L


Residuals
Absorbance
Radius
+ LSar6L


Residuals
Absorbance
Radius
+ LSar8L

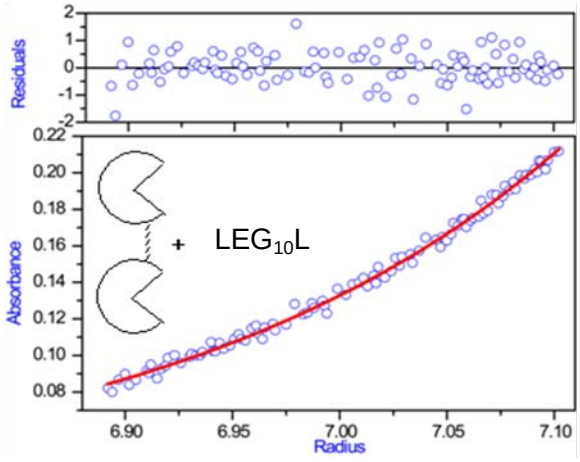

Residuals
Absorbance
Radius
+ LEG10L

Table S1. The maximum distance between the two nitrogen atoms of the sulfonamide functional groups of the bivalent ligands (extended length, Γ, Å) and the estimated molecular weights (MW) for CA, $(CA)_2$, and mixtures of $(CA)_2$ and bivalent ligands determined from sedimentation equilibrium experiments.

Proteins and complexes of dimers and bivalent ligands	**Extended length of the linker, Γ (Å)**	**Estimated MW (kDa) (95% CI)**
CA	--	34 (32, 35)
$(CA)_2$	--	58 (54, 62)
$(CA)_2$ + LEG_3L	25	57 (52, 62)
$(CA)_2$ + $LSar_0L$	31	57 (52, 62)
$(CA)_2$ + $LSar_2L$	38	56 (53, 60)
$(CA)_2$ + $LSar_4L$	46	54 (53, 55)
$(CA)_2$ + $LSar_6L$	54	58 (53, 63)
$(CA)_2$ + $LSar_8L$	62	59 (55, 63)
$(CA)_2$ + $LSar_{10}L$	69	62 (57, 68)

Derivation of Mathematical Model

This section derives the equation used to fit the fluorescence displacement data.

Figure S1. Thermodynamic scheme describing the binding of ligands to the dimer of CA. These schemes describe the definition for the binding constants

$$(1/4)\ K_{\mathrm{d}}^{\ \mathrm{inter}} = \frac{[(\mathrm{CA})_2][\mathrm{L}_2]}{[(\mathrm{CA})_2 \cdot \mathrm{L}_2^{\ \mathrm{lin}}]} \qquad \text{(S1)}$$

$$(1/2)\ K_{\mathrm{d}}^{\ \mathrm{DNSA}} = \frac{[(\mathrm{CA})_2][\mathrm{I}]}{[(\mathrm{CA})_2 \cdot \mathrm{I}]} \qquad \text{(S2)}$$

$$2\ K_{\mathrm{d}}^{\ \mathrm{intra}} = \frac{[(\mathrm{CA})_2 \cdot \mathrm{L}_2^{\ \mathrm{lin}}]}{[(\mathrm{CA})_2 \cdot \mathrm{L}_2^{\ \mathrm{cyc}}]} \qquad \text{(S3)}$$

Figure 5. Thermodynamic scheme and simulations describing the model used to analyze the fluorescence titration experiments. L2 a) The thermodynamic scheme comprises three equilibria characterized by three equilibrium constants (K_1, K_2, and K_3) and four species that contain a dimer of CA.

$$K_1 = \frac{4\,K_{\mathrm{d}}^{\mathrm{ethox}}}{K_{\mathrm{d}}^{\mathrm{inter}}} = \frac{[(\mathrm{CA})_2 \cdot \mathrm{L}_2 \cdot \mathrm{I}][\mathrm{I}]}{[(\mathrm{CA})_2 \cdot \mathrm{I}_2][\mathrm{L}_2]} \tag{S4}$$

$$K_2 = \frac{K_{\mathrm{d}}^{\mathrm{DNSA}}}{2\,K_{\mathrm{d}}^{\mathrm{intra}}} = \frac{[(\mathrm{CA})_2 \cdot \mathrm{L}_2{}^{cyc}][\mathrm{L}_2]}{[(\mathrm{CA})_2 \cdot \mathrm{L}_2 \cdot \mathrm{I}][\mathrm{I}]} \tag{S5}$$

$$K_3 = \frac{2\,K_{\mathrm{d}}^{\mathrm{intra}}}{K_{\mathrm{d}}^{\mathrm{inter}}} = \frac{[(\mathrm{CA})_2 \cdot (\mathrm{L}_2)_2]}{[(\mathrm{CA})_2 \cdot \mathrm{L}_2{}^{cyc}][\mathrm{L}_2]} \tag{S6}$$

Rewriting the concentrations of the dimer complexes in terms of the dissociation constants, statistical factors, and the concentration of $(CA)_2 \cdot L_2 \cdot I$ (eq S7-S9).

$$[(\mathrm{CA})_2 \cdot \mathrm{I}_2] = \frac{1}{4}\frac{K_{\mathrm{d}}^{\mathrm{inter}}}{K_{\mathrm{d}}^{\mathrm{DNSA}}}\frac{[\mathrm{I}]}{[\mathrm{L}_2]}[(\mathrm{CA})_2 \cdot \mathrm{L}_2 \cdot \mathrm{I}] \tag{S7}$$

$$[(\mathrm{CA})_2 \cdot (\mathrm{L}_2)_2] = \frac{K_{\mathrm{d}}^{\mathrm{DNSA}}}{K_{\mathrm{d}}^{\mathrm{inter}}}\frac{[\mathrm{L}_2]}{[\mathrm{I}]}[(\mathrm{CA})_2 \cdot \mathrm{L}_2 \cdot \mathrm{I}] \tag{S8}$$

$$[(\mathrm{CA})_2 \cdot \mathrm{L}_2{}^{cyc}] = \frac{1}{2}\frac{K_{\mathrm{d}}^{\mathrm{DNSA}}}{K_{\mathrm{d}}^{\mathrm{intra}}}\frac{1}{[\mathrm{I}]}[(\mathrm{CA})_2 \cdot \mathrm{L}_2 \cdot \mathrm{I}] \tag{S9}$$

At high concentrations of dansylamide (DNSA) employed in these experiments; the total concentration of DNSA is assumed equal to the free or unbound concentration (eq 10). The mass-balance equations for dimers and ligands (S11-S12)

$$[I]_0 = [I] \tag{S10}$$

$$[L_2]_0 = [L_2] + [(CA)_2 \cdot L_2 \cdot I] + 2[(CA)_2 \cdot (L_2)_2] + [(CA)_2 \cdot L_2{}^{cyc}] \tag{S11}$$

$$[CA_2]_0 = [(CA)_2 \cdot (I)_2] + [(CA)_2 \cdot L_2 \cdot I] + [(CA)_2 \cdot (L_2)_2] + [(CA)_2 \cdot L_2{}^{cyc}] \tag{S12}$$

Replacing the complex $(CA)_2 \cdot L_2 \cdot I$ with the variable B, for simplicity. We can express the mass-balance equation for ligand and dimer as functions of the dissociation constants, statistical factors, B, [I], and $[L_2]$ (eq 13 and 14).

$$[L_2]_0 = [L_2] + [B]\left(1 + 2\frac{K_d{}^{DNSA}}{K_d{}^{inter}}\frac{[L_2]}{[I]} + \frac{1}{2}\frac{K_d{}^{DNSA}}{K_d{}^{intra}}\frac{1}{[I]}\right) \tag{S13}$$

$$[CA_2]_0 = \left(1 + \frac{1}{4}\frac{K_d{}^{inter}}{K_d{}^{DNSA}}\frac{[I]}{[L_2]} + \frac{K_d{}^{DNSA}}{K_d{}^{inter}}\frac{[L_2]}{[I]} + \frac{1}{2}\frac{K_d{}^{DNSA}}{K_d{}^{intra}}\frac{1}{[I]}\right)[B] \tag{S14}$$

The combination of equations S10, S13, and S14 yields an expression (eq S15) for the concentration of free ligand, $[L_2]$, that depends on the total concentrations of inhibitor, $[I]_0$, ligand, $[L_2]_0$, and dimer, $[CA_2]_0$, and on the equilibrium constants of the system, $K_d{}^{inter}$, $K_d{}^{intra}$, and $K_d{}^{DNSA}$.

$$[L_2]^3 + a[L_2]^2 + b[L_2] + c = 0 \tag{S15}$$

The coefficients a, b, and c for each term of the polynomial are defined by Equations S16-S18.

$$a = \frac{1 + \frac{1}{2}\frac{K_d^{DNSA}}{K_d^{intra}}\frac{1}{[I]_0} + 2\frac{K_d^{DNSA}}{K_d^{inter}}\frac{[CA_2]_0}{[I]_0} - \frac{K_d^{DNSA}}{K_d^{inter}}\frac{[L_2]}{[I]_0}}{\frac{K_d^{DNSA}}{K_d^{inter}}\frac{1}{[I]_0}}$$

$$b = \frac{\frac{1}{4}\frac{K_d^{inter}}{K_d^{DNSA}}[I]_0 + [CA_2]_0 + \frac{1}{2}\frac{K_d^{DNSA}}{K_d^{intra}}\frac{[CA_2]_0}{[I]_0} - [L_2]_0 - \frac{1}{2}\frac{K_d^{DNSA}}{K_d^{intra}}\frac{[L_2]_0}{[I]_0}}{\frac{K_d^{DNSA}}{K_d^{inter}}\frac{1}{[I]_0}} \quad \text{(S16)}$$

(S17)

$$c = \frac{[L_2]_0([I]_0)^2}{4}\left(\frac{K_d^{inter}}{K_d^{DNSA}}\right)^2 \quad \text{(S18)}$$

There exist three solutions to a cubic equation but only one of them is physically real—it has no imaginary component (Equation S19).

$$[L_2] = -\frac{a}{3} + \sqrt[3]{R + \sqrt{Q^3 + R^2}} + \sqrt[3]{R - \sqrt{Q^3 + R^2}} \quad \text{(S19)}$$

$$Q = \frac{3b - a^2}{9}$$

$$R = \frac{9ab - 27c - 2a^3}{54}$$

The observed fluorescence signal can be described by a function of the concentration of dimers.

$$Fl_{obs} = (Fl_{max} - Fl_{min})\frac{2[(CA)_2 \cdot I_2] + [B]}{[CA_2]_0} + Fl_{min} \quad \text{(S20)}$$

Substituting the concentration of $(CA)_2 \cdot I_2$ (eq S7), and [B] (eq S14) into equation S20 yields an expression for the observed fluorescence that depends on the known total concentration of

DNSA $[I]_0$ and the varying concentration of free ligand,

$$[L_2].Fl_{obs} = \left(Fl_{\max} - Fl_{\min}\right)\frac{1+\frac{1}{2}\frac{K_d^{\ inter}}{K_d^{\ DNSA}}\frac{[I]_0}{[L_2]}}{1+\frac{1}{4}\frac{K_d^{\ inter}}{K_d^{\ DNSA}}\frac{[I]_0}{[L_2]}+\frac{K_d^{\ DNSA}}{K_d^{\ inter}}\frac{[L_2]}{[I]_0}+\frac{1}{2}\frac{K_d^{\ DNSA}}{K_d^{\ intra}}\frac{1}{[I]_0}} + Fl_{\min} \qquad \text{(S21)}$$

Substituting the value of free ligand (equation S19) in equation S21 yields an expression for fluorescence that is a function of six parameters: the total concentrations of inhibitor, $[I]_0$, ligand, $[L_2]_0$, and dimer, $[(CA)_2]_0$, and on the equilibrium constants of the system, K_d^{inter}, K_d^{intra}, and K_d^{DNSA}. The total concentrations are known, as they are chosen by the experimentalist. The dissociation constants K_d^{intra} is the sole adjustable parameter to fit the experimental curves of the observed fluorescence.

Synthesis of $(CA)_2$

Methods for the overexpression in and purification of the HCAII double mutant (K133C, C206S), mutant HCAII, from E. coli have been reported previously [1]. To a solution of mutant HCAII (11.5 mg, 0.38 μmoles) in 100 mM PBS (5.0 mL) at pH 7.2 was added a solution of bis-maleimidomethyl ether (40 μg, 0.17 μmoles) in like buffer (0.085 mL) over the course of 20 h in 8.5 μL equal portions. The volume of the crude reaction mixture was reduced to 1 mL then purified by size-exclusion chromatography using a column of Superdex 75 (16 × 600 mm) eluting with 10 mM PBS at a constant flow rate of 0.5 mL min^{-1}. The appropriate fractions were pooled to yield 8 mg (70 % yield) of $(CA)_2$. Purity and molecular weight were consistent with proposed structure as assessed by SDS-PAGE. ESI-MS m/z found 58,385 Da, calcd. 58,386 Da (apo-P_2 + $2H_2O$ from the hydration of the maleimide).

Data collection and processing

The intensity data were processed and integrated with Mosfilm [2]. Reflections were observed to 1.5 Å with measurement of 78.8% of all theoretical reflections (completeness), and a resolution cutoff was arbitrarily chosen for all further processing at 1.8 Å because the mean intensity/standard dev = 3.0 at that resolution. Statistical analysis of the intensities with a Wilson plot [CCP4 program Wilson [3]] revealed a mean isotropic temperature factor of 21.1 $Å^2$.

Molecular replacement

The structure was solved by molecular replacement, using HCAII from a published structure (PDB 1G52 [4]) as the search model. All 258 residues from the HCA search model were used, including all side chains. There was only one dominant solution found with the CCP4 program Phaser [5]. For this solution, the log likelihood gain was 870.5.

Refinement

The molecular replacement model underwent a rigid-body minimization and a Cartesian simulated annealing during which a bulk solvent correction was applied. Repeated cycles of refinement using CNS (Crystallography and NMR System) program [6] and manual rebuilding using MOLOC [7] followed to ensure that the free R-factor decreased steadily. Refinement included positional refinement, Cartesian and torsional simulated annealing, and B-factor refinement with the maximum likelihood target [8]. The model was refined against 90% of the measured data between 30.0 and 2.4 Å using the Engh and Huber stereochemical dictionary. The remaining 10% of the data were excluded from all refinement calculations to cross-validate the progress of refinement (the free R set). After 5 cycles of refinement and manual rebuilding, water molecules were automatically added to the structure provided that (a) density appeared

>2.0 ó on the 2Fo - Fc map, (b) they formed hydrogen bonds of reasonable geometry, and (c) their inclusion reduced the free R-factor. The final R-factor was 24.0 and the free R-factor was 26.9, with excellent geometry. A Ramachandran plot revealed 97% of the residues fall into allowed regions. The overall statistics for data collection and refinement are summarized in Table S2. To check for search model bias, a high-temperature electron density omit map was calculated. It revealed that the main chain positions were similar between our structure and the search model and that the side chain conformations clearly support our structure versus that of the search model. The dimer mate falls across a crystallographic 2-fold symmetry axis.

The structure of $(CA)_2$ contained one molecule of HCA in the asymmetric unit. The methylmaleimide cross-linker connected pairs of monomers; the ether oxygen of the crosslinker rested upon a crystallographic two-fold symmetry axis. Existing crystallography software does not allow bonded atoms, such as this oxygen, to lie exactly on a crystallographic axis, so bonds to the ether oxygen were modeled using fixed distance constraints. The crosslinking reaction, also, formed a diastereomeric mixture of dimers because the nucleophic attack of the thiol occurs from either face of the maleimide. The electron density for the sulfur atom, thus, appeared as a population-weighted average of the diastereomers. We modeled the sulfur atom to fit the peak of the average density of all diastereomers.

Table S2. Crystal Parameters, Data Collection, and Refinement Statistics

Data Collection	
Space Group	C2
Unit cell dimensions (a,b,c,α,β,γ)	66.7, 51.1, 80.9, 90, 107.2, 90
Number of Amino Acids, Atoms	Protein: 258, 2060, Heteroatoms: 9
Total reflections (Unique reflections)	22984
Completeness to 1.8 Å, to 1.5 Å	94.5%, 78.8%
R_{merge}	0.06
Mosaicity	1.13°
Signal to noise: Average intensity / standard dev to 1.8 Å, to 1.5 Å	10.28, 7.55
Molecular Replacement	
Log likelihood gain	870.5
R factor	37.8
Refinement	
Resolution Range / Reflections	500 – 1.8, 22984 total (20672 working set, 2312 test set)
Number of Amino Acids, Atoms	Protein: 258, 2060, Heteroatoms: 9
Ordered waters seen	201
Free R-factor (%)	24.0
R factor (%)	26.9
Structural quality statistics	
Ramachandran plot: #Residues in core, outliers regions	228, 7
Protein B-factors (mean +/- SD)	21.3 +/- 11.7
Waters B-factors (mean +/- SD)	26.06 +/- 7.5

R_{merge} for replicate reflections = $\sum \left| I_j - \langle I_j \rangle \right| / \sum \langle I_j \rangle$, where I_j is the intensity measured for reflection h and $\langle I_j \rangle$ is the average intensity for reflection h calculated from replicate data. Free R-factor is the R factor calculated only on the 10% of the reflections that were set aside for cross-validation and not used in refinement. R-factor = $\sum |Fo - Fc| / \sum |Fo|$

REFERENCES

(1) Krishnamurthy, V. M.; Semetey, V.; Bracher, P. J.; Shen, N.; Whitesides, G. M. *J. Am. Chem. Soc.* **2007**, *129*, 1312.
(2) Leslie, A. G. W. *Acta Crystallographica Section D* **2006**, *62*, 48.
(3) Bailey, S. *Acta Crystallogr D* **1994**, *50*, 760.
(4) Kim, C.-Y.; Chang, J. S.; Doyon, J. B.; Baird; Fierke, C. A.; Jain, A.; Christianson, D. W. *Journal of the American Chemical Society* **2000**, *122*, 12125.
(5) Read, R. *Acta Crystallographica Section D* **2001**, *57*, 1373.
(6) Brunger, A. T.; Adams, P. D.; Clore, G. M.; DeLano, W. L.; Gros, P.; Grosse-Kunstleve, R. W.; Jiang, J. S.; Kuszewski, J.; Nilges, M.; Pannu, N. S.; Read, R. J.; Rice, L. M.; Simonson, T.; Warren, G. L. *Acta Crystallogr D Biol Crystallogr* **1998**, *54*, 905.
(7) K. Muller, H. J. A., D.M. Doran, P.R. Gerber, K. Gubernator, and G. Schrepfer *Bull. Soc. Chim. Belg.* **1988**, *97*, 655.
(8) Adams, P. D.; Pannu, N. S.; Read, R. J.; Brunger, A. T. *Proc Natl Acad Sci U S A* **1997**, *94*, 5018.