JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Communication

pubs.acs.org/JACS

Contrast-Matched Small-Angle X-ray Scattering from a Heavy-Atom-Labeled Protein in Structure Determination: Application to a Lead-Substituted Calmodulin–Peptide Complex

Alexander Grishaev, Nicholas J. Anthis, and G. Marius Clore*

Laboratories of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892-0520, United States

Supporting Information

ABSTRACT: The information content in 1-D solution X-ray scattering profiles is generally restricted to low-resolution shape and size information that, on its own, cannot lead to unique 3-D structures of biological macromolecules comparable to all-atom models derived from X-ray crystallography or NMR spectroscopy. Here we show that contrast-matched X-ray scattering data collected on a protein incorporating specific heavy-atom labels in 65% aqueous sucrose buffer can dramatically enhance the power of conventional small- and wide-angle X-ray scattering (SAXS/WAXS) measurements. Under contrast-matching conditions the protein is effectively invisible and the main contribution to the X-ray scattering intensity arises from the heavy atoms, allowing direct extraction of pairwise distances between them. In combination with conventional aqueous SAXS/WAXS data, supplemented by NMR-derived residual dipolar couplings (RDCs) measured in a weakly aligning medium, we show that it is possible to position protein domains relative to one another within a precision of 1 Å. We demonstrate this approach with respect to the determination of domain positions in a complex between calmodulin, in which the four Ca^{2+} ions have been substituted by Pb^{2+}, and a target peptide. The uniqueness of the resulting solution is established by an exhaustive search over all models compatible with the experimental data, and could not have been achieved using aqueous SAXS and RDC data alone. Moreover, we show that the correct structural solution can be recovered using only contrast-matched SAXS and aqueous SAXS/WAXS data.

Small- and wide-angle X-ray scattering (SAXS/WAXS) in solution yield 1-D profiles that are determined by the pairwise distances between all atoms in a molecule.[1] Because of the convoluted nature of SAXS/WAXS, it is not possible, outside of the low-q range, to directly relate features of the scattering profiles to a particular structure. Further, it is generally not feasible to derive unique 3-D structures from 1-D curves, as many models may be compatible with a given scattering profile. On the other hand, direct refinement against SAXS/WAXS data in combination with other constraints, such as those from NMR data, can be extremely powerful.[2] Here we investigate the utility of scattering arising from a few heavy atoms, such as Pb^{2+}, under contrast-matched conditions at which the protein is rendered effectively invisible at low scattering angles by using a 65% aqueous sucrose solution as a solvent.[3] Contrast-matched scattering profiles arise from only a small number of heavy atoms and should therefore effectively constrain the distances between them. We demonstrate the utility of contrast-matched SAXS using, as an example, a calmodulin (CaM)–myosin light-chain kinase (MLCK) peptide complex in which the four coordinated Ca^{2+} ions have been substituted by Pb^{2+}. We show that the combination of contrast-matched SAXS, aqueous SAXS/WAXS, and NMR-derived residual dipolar couplings (RDCs)[4] is sufficient to define the positions of the two domains of CaM within a precision of 1 Å, and that the contrast-matched SAXS data are critical for discriminating among several models that are fully consistent with aqueous SAXS/WAXS and RDC data.

Examination of Ca^{2+}-loaded CaM–peptide complexes solved by NMR and crystallography reveals considerable variability in the relative positions of the N- and C-terminal domains (Figure 1). While some of this variability arises from differences in target peptide sequences leading to distinct binding modes,[5] even unique binding modes exhibit substantial structural variation. Further, none of these structures fit the SAXS/WAXS data within experimental error (Figures S4 and S5 and Table S1), and only in a handful of cases are the fits of the RDC data (measured in a

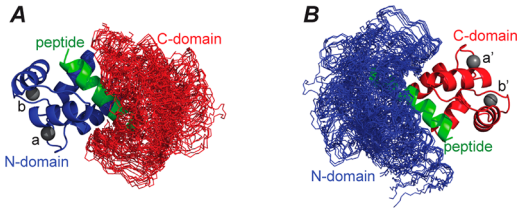

Figure 1. Distribution of the relative positions of the N- and C-domains of CaM in complexes with target peptides. Alignments on the N- (blue) and C- (red) domains are shown in (A) and (B), respectively, with the aligned domain represented by a ribbon of a single structure, and the non-aligned domain by backbone traces of the crystal structures listed in Table S1. The Ca^{2+} atoms of the N- and C-domains are shown in panels A and B, respectively, as gray spheres. The peptide from one structure (1MXE[9]) is shown as a green ribbon to illustrate how the two domains clamp the peptide.

Received: June 29, 2012
Published: August 21, 2012

dx.doi.org/10.1021/ja306359z | J. Am. Chem. Soc. 2012, 134, 14686–14689

weakly aligning medium of phage pf1[6]) to the whole complex comparable to those for the individual domains (Table S1).

Substitution of Ca^{2+} by Pb^{2+} does not significantly perturb the structure of CaM (as judged by crystallography,[7] NMR, and SAXS/WAXS), its function,[8] or its binding affinity for the MLCK peptide (see SI). Similarly, the structure of the CaM−MLCK complex appears unperturbed by the addition of sucrose as judged by NMR (see SI).

Rather than refine directly against all available SAXS/WAXS and NMR data by simulated annealing, starting from a limited set of initial coordinates,[2] we chose to evaluate the agreement with the experimental data by exhaustive sampling all possible, stereochemically feasible, relative geometries of the two domains of CaM. The latter approach guarantees that the best-fitting solution corresponds to the global minimum of the target function and is not simply one of many possible solutions, thereby removing the issue degeneracy which is universally recognized as one of the main pitfalls of biomolecular SAXS.

Fits of the RDC data to the individual domains of CaM from a large array of crystal structures indicate that the most accurate representation of the N- (residues 1−75) and C- (residues 82−148) domains in solution is that from the 1.7 Å resolution structure of a CaM−peptide complex from CaM-dependent protein kinase I (PDB code 1MXE)[9] with RDC *R*-factors of 17% (Table S1). Thus we used the N- and C-domain coordinates from the 1MXE structure throughout this study.

The rotational and translational parameters (three of each) describing the position of the two domains of CaM were systematically sampled in 2° and 1 Å steps, respectively, to generate an initial set of $\sim 4.5 \times 10^{11}$ geometries with a spatial resolution of ~1 Å. Structures were removed that (i) had steric clashes <2.5 Å between backbone or Cβ atoms of the two domains; (ii) exhibited an increase of >10% in the RDC *R*-factor (equivalent to ~10° orientational uncertainty) relative to the minimum value obtained by rigid-body minimization of the relative orientations; or (iii) had a radius of gyration (R_{gyr}) outside a range of 13−19 Å. [The experimental R_{gyr} obtained from the aqueous SAXS data via Guinier and $P(r)$ analyses is 17.8 ± 0.4 Å for the CaM−MLCK complex with Ca^{2+} and 18.3 ± 0.4 Å for the Pb^{2+}-substituted complex.] For relative domain geometries that passed the latter requirements, rotation and translation of the MLCK peptide taken from the solution structure of the CaM−MLCK complex (2BBM)[10] were obtained by minimization of a target function comprising the intermolecular NOE distance restraints from the 2BBM deposition and a repulsion term between the heavy atoms of the two CaM domains and MLCK to prevent atomic overlap. Additional geometries were then excluded in instances with steric clashes <2.5 Å between the heavy atoms of MLCK and the two domains of CaM or with NOE violations >3 Å. Finally the linker (residues 76−81) was built from a PDB database of 6250 non-redundant protein chains comprising 2.2×10^6 residues. All contiguous eight-residue stretches that did not contain either Gly or Pro (consistent with the composition of the CaM linker) were selected. The terminal residues of these eight-residue fragments were best-fitted to the backbone atoms of Lys75 of the N-domain and Glu82 of the C-domain for each CaM domain geometry. All linkers exhibiting backbone rms differences <1 Å relative to the coordinates of Lys75 and Glu82 were retained. Backbone linker geometries that did not exhibit steric clashes <2.5 Å with the two CaM domains and the MLCK peptide were processed further, and the best-fitting six-residue backbone segment corresponding to Met76−Ser81 was decorated with the appropriate side chains using residue-specific rotamers,[11] avoiding steric clashes.

The calculated SAXS/WAXS curves for the resulting 75 000 models of the CaM−MLCK complex were best-fitted to the experimental data (recorded on beamlines 12-IDC and 12-IDB, APS) using the AXES formalism which makes use of explicit water molecules to model the solvent boundary layer.[12] The scattering intensity from the Pb-substituted sample in 65% sucrose was predicted via the Debye formula from the coordinates of the Pb sites. Agreement of the aqueous SAXS/WAXS and contrast-matched SAXS data sets with structural models was evaluated using χ^2 statistics.

As is clear from Figure 2, the discriminating power of the Pb-substituted contrast-matched SAXS data is essential for selecting between candidate solutions. Three clusters of structures fit equally well to the aqueous SAXS/WAXS data (Figures 2A and 3A), and the RDCs (Figure 3C) within a few percentage points of their minimum values (Table 1). These clusters, depicted in blue (Cluster I), green (Cluster II), and red (Cluster III) in Figures 2−4 also satisfy the NOE restraints between CaM and the bound MLCK peptide (Figure 2B), indicating that all three capture the placement of the peptide within experimental error. The only observable that discriminates between the three clusters is the Pb-substituted contrast-matched SAXS data, and only one cluster, namely Cluster I, satisfies the contrast-matched SAXS data within experimental error (Figures 2A and 3B and Table 1).

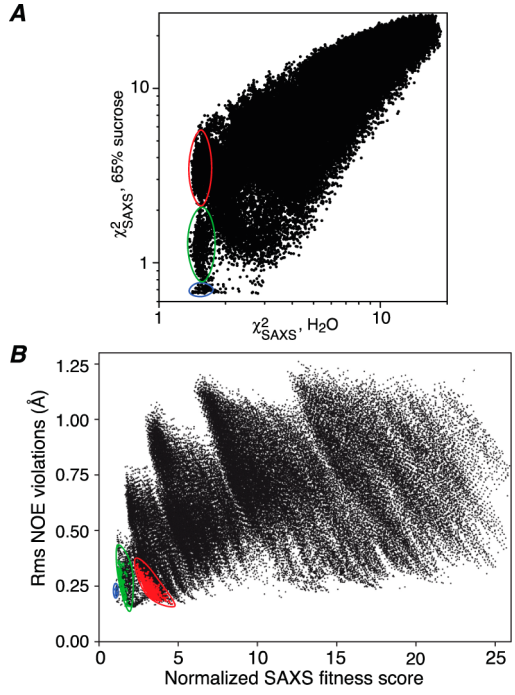

Figure 2. Agreement with SAXS data. (A) Correlation of the quality of fit of the aqueous SAXS/WAXS data with that from the contrast matched SAXS data for the Pb-substituted CaM−MLCK complex in 65% (w/v) sucrose for the 75 000 structural models generated as described in the text. (B) Correlation of the normalized SAXS fitness score with the intermolecular NOE distance restraint violations between CaM and the MLCK peptide. The three clusters discriminated by the contrast-matched SAXS data that fit equally well to the aqueous SAXS/WAXS, RDC, and NOE data are highlighted by the blue, green, and red ovals. The normalized SAXS fitness score is given by $\{[(\chi^2_{SAXS,water}/\chi^2_{SAXS,water,min})^2 + (\chi^2_{SAXS,sucrose}/\chi^2_{SAXS,sucrose,min})^2]/2\}^{1/2}$.

dx.doi.org/10.1021/ja306359z | *J. Am. Chem. Soc.* 2012, 134, 14686−14689

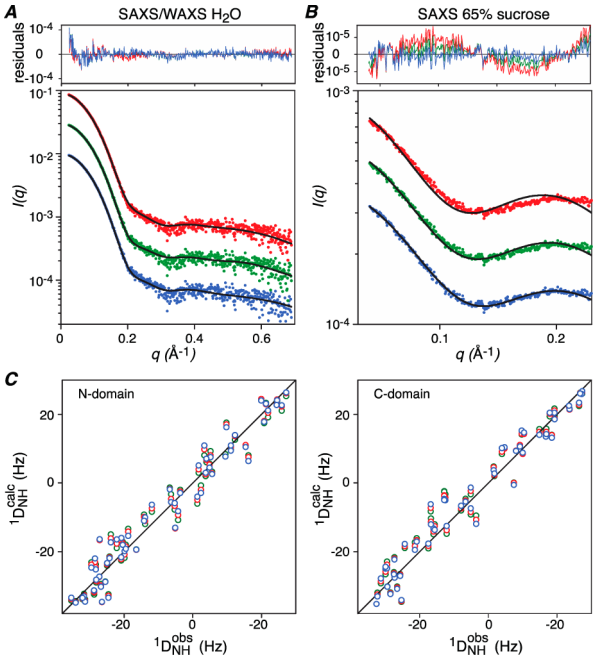

Figure 3. Agreement between observed and calculated SAXS and RDC data for the structures closest to the mean of each cluster: (A) SAXS/WAXS in water, (B) contrast-matched SAXS in 65% (w/v) sucrose, and (C) RDCs for the N- (left panel) and C- (right panel) domains. Cluster I is depicted in blue, cluster II in green, and cluster III in red.

Table 1. Parameters of the Three Clusters Selected Based on $\chi^2_{SAXS,\ water}$ Values <1.5 and Grouped According to $\chi^2_{SAXS,\ sucrose}$[a]

	Cluster I	Cluster II	Cluster III
$\chi^2_{SAXS,\ sucrose}$	0.69 ± 0.02	1.25 ± 0.22	3.42 ± 0.53
$\chi^2_{SAXS,\ water}$	1.46 ± 0.03	1.46 ± 0.03	1.46 ± 0.03
RDC *R*-factor	0.18 ± 0.00	0.19 ± 0.00	0.18 ± 0.01
D_a^{NH} (Hz)	−17.6 ± 0.1	−17.6 ± 0.2	−17.7 ± 0.1
η	0.35 ± 0.01	0.35 ± 0.01	0.35 ± 0.01
$r_{Pb(a)-Pb(a')}$ (Å)	30.9 ± 0.2	31.9 ± 0.3	33.3 ± 0.3
$r_{Pb(a)-Pb(b')}$ (Å)	34.3 ± 0.2	35.6 ± 0.9	37.4 ± 0.4
$r_{Pb(b)-Pb(a')}$ (Å)	36.9 ± 0.2	37.5 ± 0.9	38.2 ± 0.3
$r_{Pb(b)-Pb(b')}$ (Å)	39.7 ± 0.2	40.7 ± 0.3	41.8 ± 0.2

[a]RDC *R*-factor is given by $[\langle (D_{obs} - D_{calc})^2\rangle / 2\langle D_{obs}^2\rangle]^{1/2}$, where D_{obs} and D_{calc} are the observed and calculated RDCs, respectively. D_a^{NH} and η are the magnitude of the principal component of the alignment tensor and the rhombicity, respectively. There are 115 RDCs (61 and 54 for the N- and C-domains, respectively). The correspondence of the Pb^{2+} atoms within the structure is shown in Figure 4.

A structural comparison of the three clusters is shown in Figure 4, and the rms differences within and between the clusters are provided in Table 2. For Cluster I, the relative position of the two CaM domains is defined with a precision of 1 Å. When best-fitted to the N-domain, the backbone rms displacements of the C-domains of Clusters II and III relative to Cluster I are much larger, 1.8 and 5.4 Å, respectively, well outside the coordinate precision of Cluster I. Structural differences between the three clusters reflect systematic lengthening of the Pb−Pb distances from Cluster I to Clusters II and III (Figure 4 and Table 1).

Given the discriminating power of the Pb-substituted contrast-matched SAXS data, can such data be directly used to accurately extract Pb−Pb distances? To assess this we carried out fits to the contrast-matched SAXS data using a Monte Carlo-generated random sampling of 4-atom geometries to represent the Pb sites

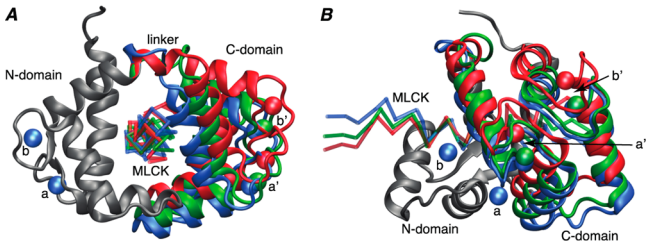

Figure 4. Structural comparison of Clusters I, II, and III. The N-domain (gray) of the three clusters is superimposed, illustrating the relative displacements of the C-domain of Clusters I (blue), II (green), and III (red). The backbone of CaM is displayed as a ribbon diagram, the backbone of the MLCK peptide is shown as a Cα trace, and the positions of the Pb^{2+} ions are displayed as spheres. Only Cluster I accounts for the Pb-substituted contrast-matched SAXS data. The backbone rms displacement of the C-domain of Cluster I relative to the 1MXE coordinates[9] is 2.5 Å.

Table 2. Backbone Root-Mean-Square Difference between the C-Terminal Domains of Clusters I, II, and III, When Best-Fitted to the N-Domain[a]

	backbone rms difference (Å)		
Cluster	I	II	III
I	1.1	1.8	5.4
II		2.4	3.9
III			1.4

[a]Diagonal elements are the average rms difference within a cluster to the structure closest to the mean coordinate positions of the cluster, and off-diagonal elements are rms differences between the structures closest to the means of the respective clusters.

Table 3. Pb−Pb Distances Determined Directly from the Pb-Substituted Contrast-Matched SAXS Data

no.[a]	Pb−Pb (Å)
4[b]	31.5 ± 1.0, 33.8 ± 0.9, 36.5 ± 1.1, 39.6 ± 1.1
5[c]	11.1 ± 2.9, 30.2 ± 2.3, 33.9 ± 1.3, 36.3 ± 1.3, 40.6 ± 1.5
6	6.4 ± 3.6, 14.7 ± 3.1, 29.9 ± 1.9, 32.3 ± 1.9, 38.1 ± 1.8, 41.1 ± 1.3

[a]Number of unknown distances. [b]Only the four interdomain Pb−Pb distances are varied, and the two intradomain Pb−Pb distances are fixed to their known value of 11.7 Å. [c]The two intradomain Pb−Pb distances are varied but constrained to be equal.

(see SI). We consider three cases: four variable interdomain Pb−Pb distances with the two intradomain distances fixed to the values measured from the domain coordinates (11.7 Å); five variables comprising the four interdomain Pb−Pb distances with the two intradomain distances constrained to have the same value; and six variables in which all four interdomain and two intradomain Pb−Pb distances are allowed to vary. For the four-variable case, the average interdomain Pb−Pb distances derived from the contrast-matched SAXS data have uncertainties of only ~1 Å (Table 3) and are in excellent agreement with the corresponding Pb−Pb distances in Cluster I (cf. Table 1). The five-variable case results in larger uncertainties, but agreement with the Cluster I Pb−Pb distances is still excellent and within the uncertainties of the distance estimates. Finally, the six-variable case results in reasonable agreement with the longer, interdomain Pb−Pb distances from the all-atom model but exhibits marginal agreement with the correct intradomain Pb−Pb separations. These calculations therefore indicate that up to five (from a total of six) Pb−Pb distances can be extracted both

dx.doi.org/10.1021/ja306359z | J. Am. Chem. Soc. 2012, 134, 14686−14689

precisely and accurately from a single curve of the contrast-matched SAXS data arising from four Pb labels. Accelerated deterioration of the uncertainties of the shorter separations is due to increased uncertainty at wider angles and the limited q_{max} of the fitted data, as the impact of the shorter Pb−Pb distances is increasingly felt at higher q.

In light of the above results, we sought to investigate whether the contrast-matched SAXS and aqueous SAXS/WAXS data alone are sufficient to arrive at the correct solution without recourse to filtering by RDCs and intermolecular NOE distance violations. Using the same grid procedure, structures were filtered by the fits to the X-ray scattering data using a cutoff of $\chi^2_{SAXS,water} < 1.69$ and $\chi^2_{SAXS,sucrose} < 0.72$, the absence of steric clashes, an R_{gyr} of 13−19 Å, and the ability to form a linkage between the N- and C-domains. The results in Figure 5 indicate that the cluster with the lowest normalized SAXS fitness score yields solutions with a coordinate accuracy of better than 1 Å by reference to Cluster I.

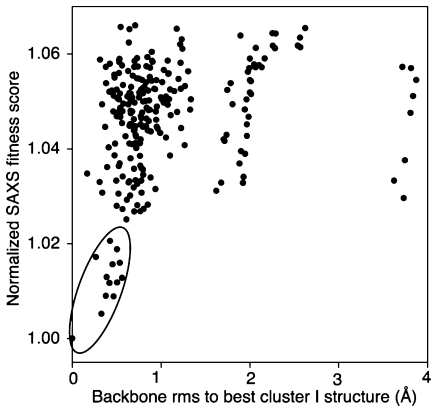

Figure 5. Accuracy of best-fitting models ($\chi^2_{SAXS,water} < 1.69$ and $\chi^2_{SAXS,sucrose} < 0.72$) obtained from the grid search procedure using aqueous and contrast-matched SAXS data as the only experimental restraints. The normalized SAXS fitness score is defined in the legend to Figure 2. The best Cluster I structure is the structure closest to the mean coordinate positions of the Cluster I structures.

While this work capitalizes on the ability of CaM to specifically bind Pb^{2+} in place of Ca^{2+}, applications of this approach can be readily extended beyond metal-binding proteins by incorporating heavy-atom ions such as Pb^{2+} or Hg^{2+} into EDTA moieties conjugated via disulfide bonds to engineered surface cysteines as routinely done in NMR paramagnetic relaxation enhancement studies.[13] Although the EDTA−metal moiety samples a large region of conformational space, the metal−metal separations measured by contrast-matched SAXS are simple linear averages of all the conformations present in solution and each metal atom can therefore be represented by a single average position.

Although, in principle, similar information has been obtained from small-angle neutron scattering (SANS) of ^{240}Pu,[14] or X-ray scattering of DNA with attached gold nanoclusters,[15] the present approach does not suffer from the lower signal-to-noise of SANS and the use of a radioactive isotope, or the necessity to decompose the observed data into individual scattering functions from measurements on a series of samples.

Finally, the results obtained without recourse to filtering by NMR data suggest that accurate "triangulation-driven" assembly of multi-component protein architectures, based only on a combination of aqueous and contrast-matched/heavy-atom-labeled SAXS data, is feasible providing the structures of the individual subunits are known.

■ ASSOCIATED CONTENT

Supporting Information

Experimental procedures, data tables, figures, and complete ref 1b. This material is available free of charge via the Internet at http://pubs.acs.org. Coordinates of the best-fitting Cluster I model and experimental data (PDB 2LV6) have been deposited in the Protein Data Bank (http://www.rcsb.org).

■ AUTHOR INFORMATION

Corresponding Author

mariusc@mail.nih.gov

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank A. Bax for useful discussions, and S. Seifert and X. Zuo for assistance with SAXS data collection. This work was supported by the intramural program of NIDDK and the Intramural AIDS Antiviral Program of the Office of the Director of the NIH (G.M.C.). The shared beamline resource (PUP-77 agreement between NCI, NIH, and Argonne National Laboratory) at the APS is acknowledged.

■ REFERENCES

(1) (a) Koch, M. H.; Vachette, P.; Svergun, D. I. *Q. Rev. Biophys.* **2003**, *36*, 147. (b) Hura, G. L.; et al. *Nature Methods* **2009**, *6*, 606.

(2) (a) Grishaev, A.; Wu, J.; Trewhella, J.; Bax, A. *J. Am. Chem. Soc.* **2005**, *127*, 16621. (b) Schwieters, C. D.; Clore, G. M. *Biochemistry* **2007**, *46*, 1152. (c) Grishaev, A.; Tugarinov, V.; Kay, L. E.; Trewhella, J.; Bax, A. *J. Biomol. NMR* **2008**, *40*, 95. (d) Schwieters, C. D.; Suh, J. Y.; Grishaev, A.; Ghirlando, R.; Takayama, Y.; Clore, G. M. *J. Am. Chem. Soc.* **2010**, *132*, 13026. (e) Takayama, Y.; Schwieters, C. D.; Grishaev, A.; Ghirlando, R.; Clore, G. M. *J. Am. Chem. Soc.* **2011**, *133*, 424.

(3) Lipfert, J.; Doniach, S. *Annu. Rev. Biophys. Biophys. Chem.* **2007**, *36*, 307.

(4) Bax, A.; Grishaev, A. *Curr. Opin. Struct. Biol.* **2005**, *15*, 563.

(5) (a) Hoeflich, K. P.; Ikura, M. *Cell* **2002**, *108*, 739. (b) Maximciuc, A. A.; Putkey, J. A.; Shamoo, Y.; Mackenzie, K. R. *Structure* **2006**, *14*, 1547.

(6) Clore, G. M.; Starich, M. R.; Gronenborn, A. M. *J. Am. Chem. Soc.* **1998**, *120*, 10571.

(7) Kursula, P.; Majava, V. *Acta Crystallogr.* **2007**, *F63*, 653.

(8) Fullmer, C. S.; Edelstein, S.; Wasserman, R. H. *J. Biol. Chem.* **1985**, *260*, 6816.

(9) Clapperton, J. A.; Martin, S. R.; Smerdon, S. J.; Gamblin, S. J.; Bayley, P. M. *Biochemistry* **2002**, *41*, 14669.

(10) Ikura, M.; Clore, G. M.; Gronenborn, A. M.; Zhu, G.; Klee, C. B.; Bax, A. *Science* **1992**, *256*, 632.

(11) Krivov, G. G.; Shapovalov, M. V.; Dunbrack, R. L., Jr. *Proteins* **2009**, *77*, 778.

(12) Grishaev, A.; Guo, L.; Irving, T.; Bax, A. *J. Am. Chem. Soc.* **2010**, *132*, 15484.

(13) Clore, G. M.; Iwahara, J. *Chem. Rev.* **2009**, *109*, 4108.

(14) Seeger, P. A.; Rokop, S. E.; Palmer, P. D.; Henderson, S. J.; Hobart, D. E.; Trewhella, J. *J. Am. Chem. Soc.* **1997**, *119*, 5118.

(15) Mathew-Fenn, R. S.; Das, R.; Silverman, J. A.; Walker, P.; Harbury, P. A. B. *PloS ONE* **2008**, *3*, e3229.