Use of High Pressure NMR Spectroscopy to Rapidly Identify Proteins with Internal Ligand-Binding Voids

Donald Gagné[1,†], Roksana Azad[1,2], Uthama R. Edupuganti[1,2], Justin Williams[1,‡], James M. Aramini[1], Kazuyuki Akasaka[3], Kevin H. Gardner[1,4,5,*]

[1]Structural Biology Initiative, CUNY Advanced Science Research Center, New York, NY 10031, USA

[2]Ph.D. Program in Biochemistry, Graduate Center, City University of New York, NY 10016, USA

[3]Kyoto Prefectural University of Medicine, Kyoto 602-8566, Japan

[4]Department of Chemistry and Biochemistry, City College of New York, NY 10031, USA

[5]Ph.D. Programs in Biochemistry, Chemistry and Biology Ph.D. Programs, Graduate Center, City University of New York, NY 10016, USA

†: current address: KisoJi Biotechnology, Montréal, QC, Canada

‡: current address: Department of Physiology and Cellular Biophysics, Columbia University Irving Medical Center, New York, NY 10032

[*]: address correspondence to:

Kevin Gardner, Structural Biology Initiative, CUNY Advanced Science Research Center, 85 St. Nicholas Terrace, New York, NY 10031 USA, E-mail: kgardner@gc.cuny.edu; phone: +1 212 413 3220

Author ORCIDs:
Donald Gagné: 0000-0002-3938-6567
Roksana Azad: 0000-0001-9909-2889
Uthama R. Edupuganti: 0000-0003-0571-4101
Justin Williams: 0000-0003-4108-5808
James M. Aramini: 0000-0001-8926-4210
Kazuyuki Akasaka: 0000-0002-9611-9086
Kevin H. Gardner: 0000-0002-8671-2556

Classification: Biological Sciences / Biophysics and Computational Biology

Keywords: high pressure NMR, ligand binding, protein cavities

Abstract

Small molecule binding within internal cavities provides a way to control protein function and structure, as exhibited in numerous natural and artificial settings. Unfortunately, most ways to identify suitable cavities require high-resolution structures *a priori* and may miss potential cryptic sites. Here we address this limitation via high-pressure solution NMR spectroscopy, taking advantage of the distinctive nonlinear pressure-induced chemical shift changes observed in proteins containing internal cavities and voids. We developed a method to rapidly characterize such nonlinearity among backbone ^{1}H and ^{15}N amide signals without needing to have sequence-specific chemical shift assignments, taking advantage of routinely available ^{15}N-labeled samples, instrumentation, and 2D ^{1}H/^{15}N HSQC experiments. From such data, we find a strong correlation in the site-to-site variability in such nonlinearity with the total void volume within proteins, providing insights useful for prioritizing domains for ligand binding and indicating mode-of-action among such protein/ligand systems. We suggest that this approach provides a rapid and useful way to rapidly assess otherwise hidden dynamic architectures of protein that reflect fundamental properties associated with ligand binding and control.

19 **Significance Statement**

20 Many proteins can be regulated by internally binding small molecule ligands, but
21 it is often not clear *a priori* which proteins are controllable in such a way. Here we
22 describe a rapid method to address this challenge, using solution NMR spectroscopy to
23 monitor the response of proteins to the application of high pressure. While the locations
24 of NMR signals from most proteins respond to high pressure with linear chemical shift
25 changes, proteins containing internal cavities that can bind small molecule ligands
26 respond with easily identified non-linear changes. We demonstrate this approach on
27 several proteins and protein/ligand complexes, suggesting that it has general utility.

Introduction

Small molecule cofactors and ligands play critical roles in controlling protein structure and function, the understanding of which often gives insight both into natural and artificial modes of regulation. Of particular interest are identifying sites where the irregular structures of proteins give rise to cavities, voids and other features which can serve as internal sites for such compounds to bind (**Fig. 1**) and allosterically control protein structure and function (1, 6-9). The traditional method to identify such locations – high resolution X-ray crystallography, ideally with better than 2 Å resolution to aid the identification of internally bound waters – can be powerful, particularly when combined with computational analyses for cavity identification (5, 10-16) or experimentally solving multiple structures of proteins soaked with different organic solvents or small molecule fragments (17, 18). Accordingly, this approach relies on having well-diffracting crystals, which are not always available for all systems and can be time-consuming to produce even when successful.

As an alternative method to experimentally determine which proteins might contain internal cavities suitable for ligand binding, we explored the potential for using high pressure solution NMR to do so. We thought this approach might be useful given the need for proteins to undergo dynamic changes to allow for ligand binding within pre-existing internal cavities and voids (**Fig. 1**). While such changes may occur rarely at ambient pressure, elevated pressure will easily increase the equilibrium populations of the less populated, low-volume conformers associated with hydration of cavities and voids (20). Such low-lying excited state conformers (N'), if present, will usually equilibrate with the ground state folded conformer (N) rapidly on the NMR time scale

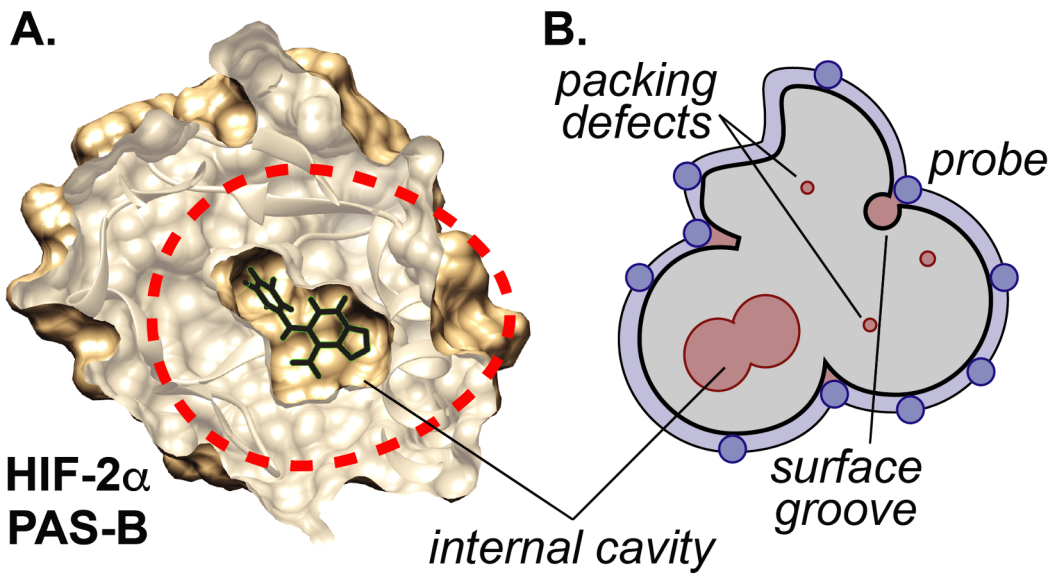

Fig. 1: Example of internal ligand binding cavity and computational analysis. A. Example of protein/ligand complex utilizing an internal cavity (HIF-2α PAS-B, PDB: 3f1o (6)) that is sequestered over 6 Å from solvent. In the apo- form of this protein (not shown), a pre-formed 290 $Å^3$ cavity with 8 crystallographically-ordered water molecules is present at this site. B. Schematic definitions of cavities, voids, and algorithm used by *ProteinVolume* (5) as used for Fig. 3 and onwards in the manuscript. We refer to *cavities* as internal openings larger than a single water molecule (V > 30 $Å^3$ (19)), while *voids* are more generally non-protein filled spaces that include cavities along with other types of packing defects distributed throughout and around a protein. With a protein structure available (gray), the *total void volume* can be straightforwardly calculated by the difference of the solvent-excluded volume of a protein (generated by rolling probes over the molecular surface, typically with water-sized radii) and the volume taken by protein atoms = volume under blue surface minus gray volume (5).

(τ<<ms), giving rise to averaged single peaks in multidimensional spectra which exhibit non-linear chemical shift changes as pressure affects the N ⇌ N' equilibrium (21). Based on prior work by several groups (21-24), we anticipated that this effect might be easily detected by solution NMR at pressures in the 1000-2000 bar range, below the pressures which typically partially or completely unfold proteins (21, 25, 26).

To evaluate this approach, we examined the effects of high hydrostatic pressure on the NMR chemical shifts of a collection of protein domains and protein/ligand

58 complexes. Many of these proteins are members of the Per-ARNT-Sim (PAS) family of
59 ligand-controlled protein/protein interaction domains, which often internally bind different
60 small molecule cofactors to sensitize them to environmental factors like O_2, light, and
61 xenobiotics (27, 28). Changes to the occupancies or configurations of these cofactors
62 trigger conformational changes in the surrounding protein, regulating the activity of
63 various effector domains in natural and engineered proteins (28, 29). While a number
64 of high-resolution structures of apo- and ligand-bound PAS domains have been solved,
65 these models and predictions which can be made from them can provide insights into
66 only a small fraction of the many thousands of "orphan" PAS domains without known
67 ligands. Complementing these proteins, we added additional proteins and
68 protein/ligand complexes from a wide range of domain types – including some which
69 are known to bind ligands internally, some not – to establish the generality of this
70 approach for quickly probing protein structure and function.

71 Here we test the ability of high-pressure NMR to rapidly identify void-containing
72 proteins, with three key advances. First, by analyzing pressure titration data from over
73 40 proteins and protein/ligand complexes, we show that an easily accessible metric –
74 how differently sites within a protein respond to increasing pressure, as assessed by the
75 diversity of non-linear chemical shift perturbations observed in a simple titration without
76 requiring site-specific assignments – correlates well with the void volume within a
77 protein. We find that this metric is robust enough to predict total void volume on its own,
78 allowing the prioritization of potential ligand-binding capability amongst several targets.
79 Second, we demonstrate that internal ligand binding within a protein can reduce this
80 heterogeneity, quickly providing information on ligand binding and, in certain cases,

81 mode of action. Finally, we illustrate how this method can also be used to rapidly
82 assess the impact of point mutations/repacking, such as those used to fill or generate
83 cavities to facilitate artificial control, on the prevention or enabling of water entry. Taken
84 together, our data show the general utility of these rapidly acquired, easily analyzed
85 data to provide this important biophysical characterization of new proteins.
86

87 **Results**

88 *Evaluation of non-linear NMR chemical shift responses to high pressure:* We began by
89 examining the pressure dependence of NMR signals from a set of well-characterized
90 proteins using the workflow in **Fig. 2**. For each U-^{15}N labeled protein sample, we
91 acquired ^{1}H/^{15}N HSQC spectra at increasingly higher pressures from 20-2500 bar. After
92 each individual dataset was acquired at high pressure, we lowered the pressure to 20
93 bar and acquired a spectrum to confirm the reversibility of conformational changes; after
94 any sign of irreversible changes in peak intensity or location, the pressure series was
95 stopped. These series were typically composed of 21 spectra each taking 60 min
96 apiece, for a total of approximately 21 hr. Post-processing, peaks were picked and the
97 pressure dependence of changes of their ^{1}H and ^{15}N chemical shifts were
98 independently fit to a second-order polynomial equation:
99

100 $$\delta_i = a_i + b_i p + c_i p^2 \quad \textbf{(Eq. 1)}$$

101

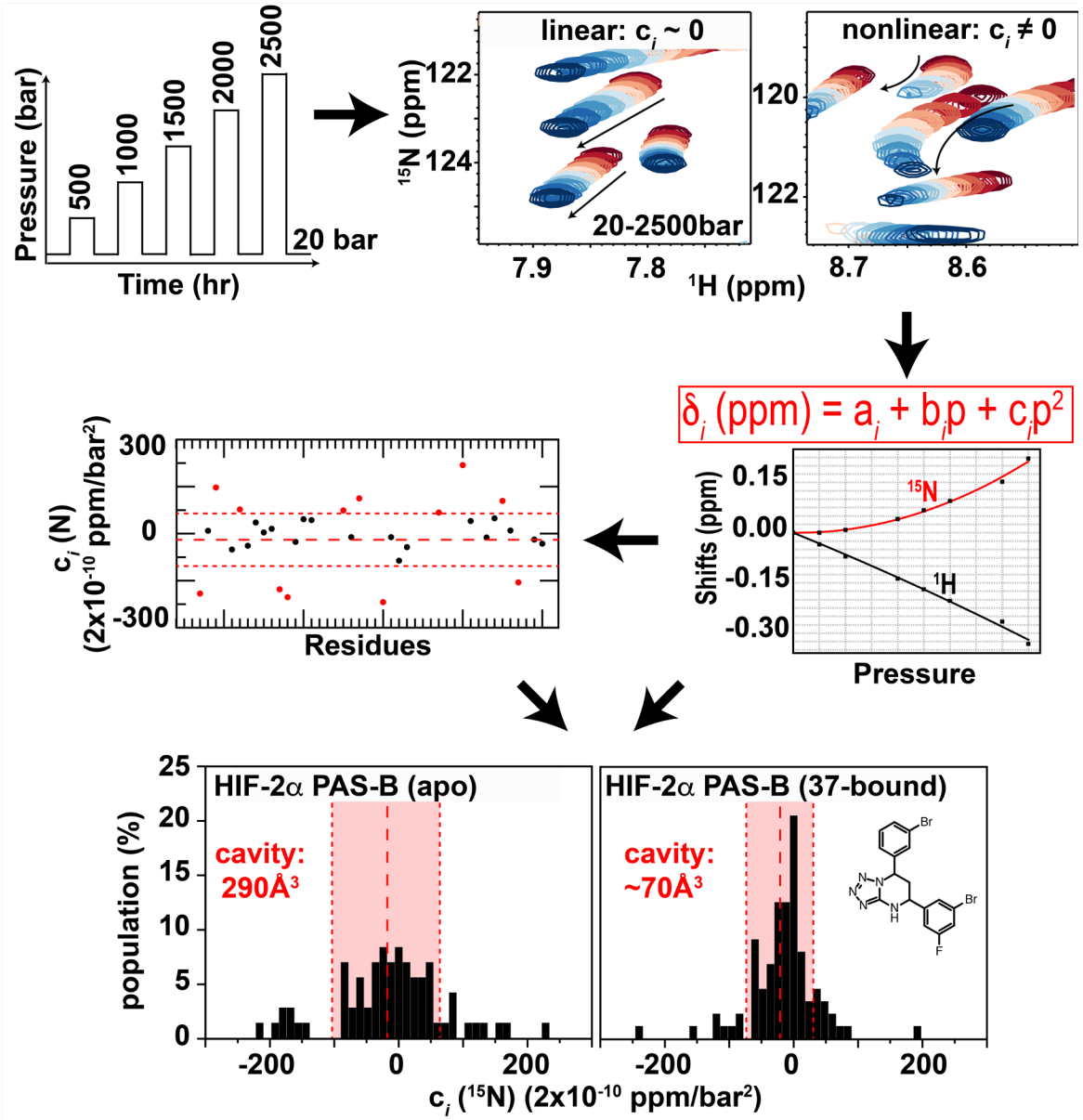

Fig. 2: Workflow of pressure NMR analyses. For each protein analyzed, ^{1}H/^{15}N HSQC spectra were acquired at increasing pressures from 20-2500 bar, interleaving additional spectra at 20 bar between those at higher pressure to assess protein reversibility. Following processing, peaks were picked and chemical shifts monitored as a function of increasing pressure. Independently treating ^{1}H and ^{15}N movements, the pressure-dependent chemical shift changes of each peak are fit to the second order polynomial indicated to obtain linear (b_i) and nonlinear (c_i) coefficients. Subsequent analyses of the nonlinear (c_i) coefficients include either plotting c_i values as a function of residue number (for proteins with backbone chemical shift assignments) to identify regions with likely pressure-dependent conformational changes or to simply generate histograms of c_i values to give a quick initial characterization of likely ability to adopt multiple folded conformations.

102

103 As established by Akasaka and co-workers (21, 25), the linear (b_i) and nonlinear
104 (c_i) coefficients of these pressure responses reflect different properties of each protein.
105 To evaluate various analyses of these data, we examined an initial group of nine
106 proteins ("test set") with high resolution structures with total void volumes ranging from
107 approximately 1500-8500 $Å^3$ as assessed by ProteinVolume (5). This approach
108 provides a sum of the volumes of a wide array of packing defects, cavities, etc.
109 regardless of their size or distribution within a protein structure by simply calculating the
110 difference between the solvent-accessible volume of a protein and the volume occupied
111 by protein atoms (**Fig. 1B**). Pressure titration data were additionally recorded from
112 more than 30 additional proteins and protein ligand complexes, with varying degrees of
113 structural information to build the "complete set" of data for subsequent analyses.
114 As an initial analysis, we examined the absolute values of the two chemical shift
115 coefficients ($|b_i|$, $|c_i|$), separately averaged over all backbone amide protons and
116 nitrogens (typically 25-125) within each protein. From these analyses, we confirmed
117 that the averaged ^1H and ^{15}N $|b_i|$ values were fairly uniform across proteins with less
118 than two-fold variation, while the corresponding $|c_i|$ values varied over 4-10 fold ranges
119 (**Fig. S1**). These data expand the original data sets of Akasaka and Li (25) from seven
120 proteins and complexes to over 40 total, add an independent measurement and
121 analysis of the GB1 protein to assess the impact of different equipment, labs, and
122 software for data acquisition and analyses (*e.g*. 750 MHz in Kobe University by K.A.;
123 700-800 MHz in New York by K.H.G.), and support the original interpretations that the
124 linear $|b_i|$ component is relatively fixed among systems while $|c_i|$ depends on a protein-
125 specific feature (25).

126 Further investigating these pressure-dependent chemical shift changes, we
127 noted a trend towards proteins with larger total void volumes having both higher c_i
128 values (**Fig. S1**) and a greater range of individual residue-specific c_i parameters than
129 proteins with smaller voids, both of which we thought might reflect a more heterogenous
130 structural response. While the larger c_i values are captured by the previously-described
131 average $|c_i|$ parameter (25), we examined several ways to quantitate the heterogeneity
132 in c_i values, settling on a combination of histograms and the standard deviation of the c_i
133 (stdev(c_i)). While the correlation between average $|c_i|$ and stdev(c_i) values for amide ^{1}H
134 and ^{15}N shifts (**Fig. S2**) suggest similarities between the two metrics, we opted to
135 proceed using stdev(c_i) to take advantage of the larger range of values provided by
136 dropping the absolute value operation. We also observed a high correlation between
137 the stdev(c_i) values of ^{1}H and ^{15}N (**Fig. S3**), suggesting that non-linear shift changes of
138 either nucleus report on pressure-induced changes despite differences in the structural
139 factors which influence them (30). Our subsequent analyses utilized stdev(c_i) of ^{15}N
140 chemical shift changes (= stdev(c_i [^{15}N])), which are thought to be most strongly
141 influenced by changes in backbone torsion angles (25, 30).
142 To examine the linkage of stdev(c_i) to total void volume, we measured the
143 nonlinear components (c_i [^{15}N]) of a test set of nine proteins (with one, GB1, repeated in
144 duplicate) with known structures and a range of total void volumes (**Fig. 3**). Histograms
145 of stdev(c_i) parameters exhibited the previously-mentioned variability, with proteins
146 having larger total void volumes typically having broader distributions than those with
147 smaller total void volumes (**Fig 3 inset**). We interpreted this trend as stemming from
148 voids enabling proteins to increasingly shift from the native N conformation to a second

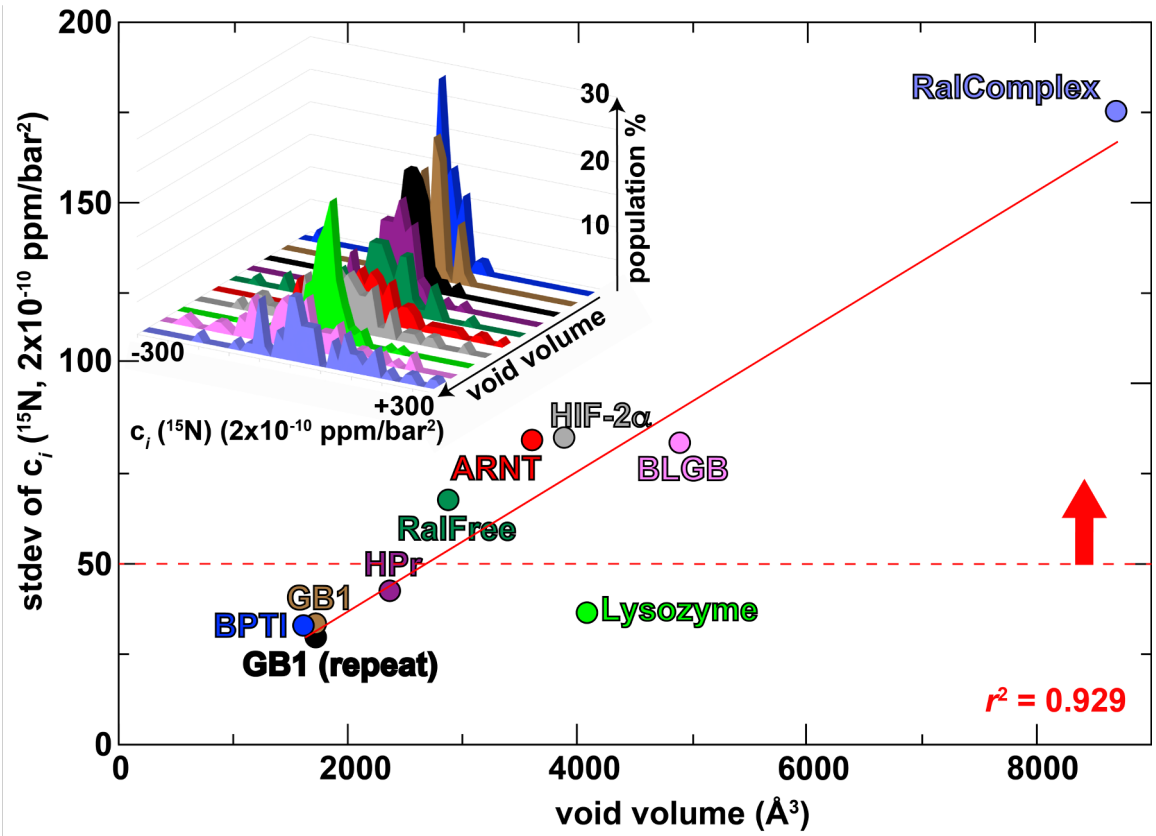

Fig. 3: **Increasing diversity of non-linear pressure dependent chemical shift changes correlates with increased void volume**. Total void volumes calculated by ProteinVolume (5) for each proteins are presented on the x-axis; the y-axis plots the standard deviation of the non-linear c_i parameters for the amide ^{15}N chemical shifts of each protein, typically using data from 25-125 crosspeaks for each protein. The linear regression (red line, y= –1.564+0.0194x, omitting the lysozyme point) shows a correlation coefficient (r^2) of 0.929. The red dashed line indicates an arbitrary value of approximately 50 (x2x10^{-10}) ppm/bar^2 as suggesting the potential for small molecule binding, as noted in the text. Inset: histograms of each of the ten proteins showing the distributions of measured ^{15}N c_i parameters for the same set of proteins (as indicated by the same color scheme).

149 folded N' conformation under pressure, reflected in the non-linear chemical shift
150 responses as the N' is progressively populated. Quantitating the breadth of these
151 distributions by stdev(c_i [^{15}N]) and plotting these versus total void volume showed a
152 linear correlation coefficient (r^2) of 0.929, supporting the linkage between pressure-
153 induced chemical shift nonlinearities to identify cavities. Of note, the proteins with a
154 stdev(c_i [^{15}N]) above 50 (x 2x10^{-10} ppm/bar^2) are all known to bind other proteins or

155 small molecules, suggesting an arbitrary value which could be useful to prioritize for
156 ligand-screening.

157 We note that the primary outlier of the observed linear correlation, hen egg white
158 lysozyme, is quite thermostable with a T_m of almost 75°C (31), likely hampering its
159 transition into an excited state under pressure unless its intrinsic stability of its basic
160 folded state is substantially lowered (*e.g.* by cooling close to the cold denaturation
161 temperature (32)). In addition, we cannot rule out potential contributions from lysozyme
162 being an enzyme instead of a signal transduction component, which may contribute a
163 different ability to adopt alternative conformations reflected in pressure-induced non-
164 linear chemical shift changes. This remains to be seen in future studies.

165

166 *Use of pressure NMR to reveal ligand binding mode of action – Well-defined cases:*
167 Our correlation between increased heterogeneity of pressure-induced chemical shift
168 changes and void size makes a strong prediction that this route should provide a rapid
169 way to assess ligand binding: ligands which bind within the protein and reduce total void
170 volume should decrease the stdev(c_i) value of spectra recorded on the receptor. To test
171 this prediction, we evaluated how ligand binding affects the PAS-B domains of the
172 human HIF-2α and ARNT proteins, both of which are involved in the human hypoxia
173 response (33) and contain internal cavities known to bind artificial small molecule
174 ligands (1, 3, 6).

175 Prior studies of the HIF-2α PAS-B domain by our group and others (1, 6, 34, 35)
176 have shown that it contains a single internal water-filled 290 Å^3 cavity with no obvious
177 access to external solvent. However, a variety of screening efforts – from NMR-based

178 fragment binding screens to high throughput screens of HIF-2α/ARNT disruption (1, 2,
179 6, 35) or protein stability (35) – have identified a wide range of small molecules which
180 bind into this cavity with nano- to micromolar affinities. High-resolution X-ray structures
181 of these complexes show that the ligands displace water and reduce HIF-2 activity by
182 impacting HIF-2α/ARNT interactions (36-38). To accommodate these binding events
183 within a solvent-inaccessible cavity, the HIF-2α PAS-B protein must dynamically
184 fluctuate to allow for a conformation which allows the entry of small molecules into the
185 interior (39). Our pressure NMR analysis confirmed this hypothesis, showing a very
186 broad distribution of c_i (^{15}N) responses with a correspondingly large stdev(c_i [^{15}N]) of 85
187 x ($2x10^{-10}$ ppm/bar^2) (**Figs. 4A and S4**). We anticipated that repeating these
188 measurements in the presence of two nanomolar-affinity compounds (**2** (1) and **37** (2))
189 would show smaller non-linear chemical shift changes than the apo protein given the
190 smaller total void volume and expected reduced flexibility of the protein/ligand
191 complexes. Our data supported this, as we observed decreases in stdev(c_i [^{15}N]) from
192 85 x ($2x10^{-10}$ ppm/bar^2) for the apo protein to 64 (**2**-bound) and 53 (**37**-bound) x ($2x10^{-10}$
193 ppm/bar^2), respectively. Notably, these decreases do not solely correlate with loss of
194 void volume in some cases: measurements conducted on HIF-2α D1, a
195 computationally-repacked variant with five point mutations which reduce the cavity
196 volume to 77 Å^3 (40) while retaining function, shows an *increase* in stdev(c_i [^{15}N]) up to

197 131 x ($2x10^{-10}$ ppm/bar^2) (**Fig. 4A**). We speculate that differences between some

198 aspect of this engineered protein, perhaps in the dynamics or thermodynamic stability,

199 and natural proteins contribute to non-linear responses to pressure.

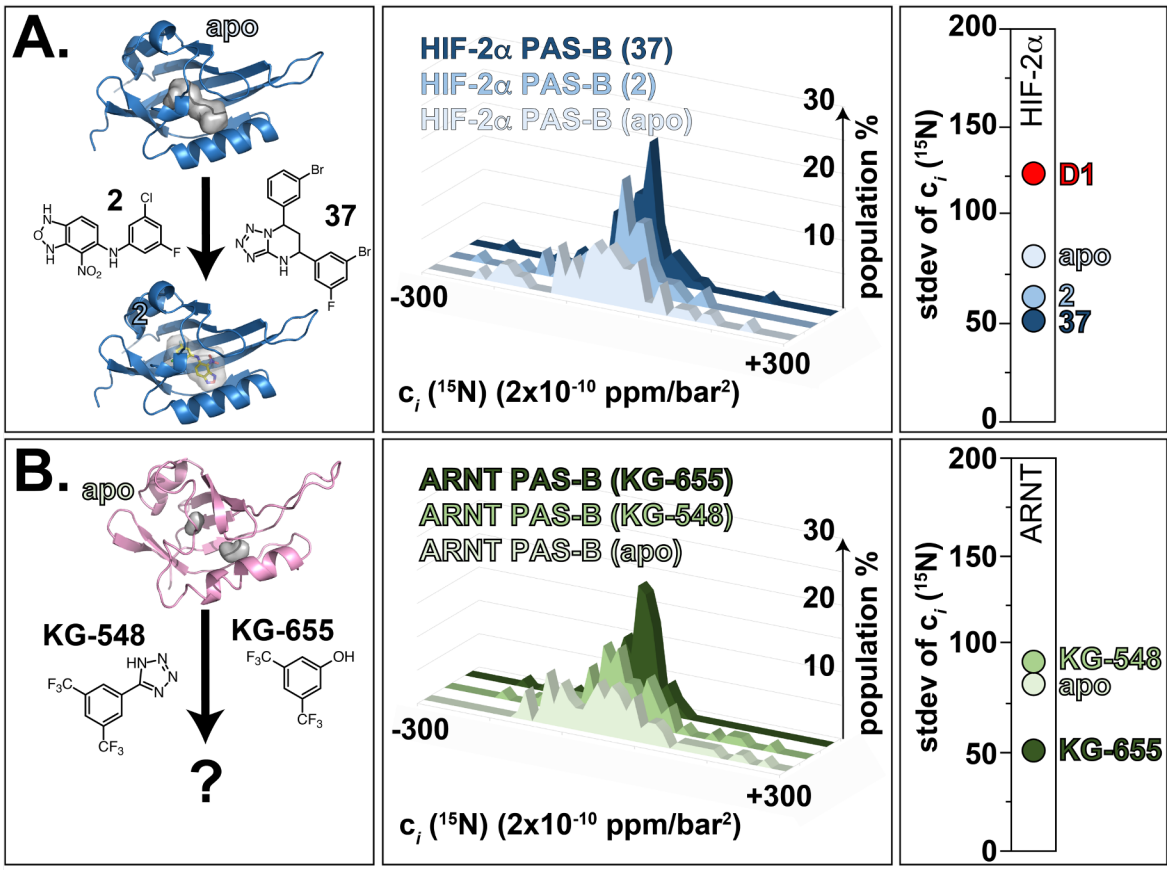

Fig. 4: High pressure NMR can provide mode-of-action information on ligand binding, even in the absence of structural information. (Left panel) A. Three-dimensional ribbon diagrams of HIF-2α PAS-B in its apo (PDB: 3f1p) and holo forms (complexed with **2**, PDB: 4ghi), highlighting the location of an internal cavity for ligand binding. Chemical structures of two high-affinity inhibitors, **2** and **37** (1, 2) are also shown. B. Three-dimensional ribbon diagram of ARNT PAS-B (PDB: 3f1p) highlighting the location of an internal cavities, along with the structures of two moderate affinity binders, **KG-548** and **KG-655** (3). No structural details are available for the bound form. (Central panel) Histograms of the ^{15}N c_i coefficients measured on apo- and holo forms of HIF-2α and ARNT PAS-B domains using the approach diagrammed in **Fig. 2**. (Right panel) Comparisons of the stdev(c_i[^{15}N]) values of HIF-2α and ARNT PAS-B apo and ligand-bound samples, showing substantial effects as cavities are occupied by water (apo forms), small molecule ligands or protein sidechains.

200 We used the same approach with the ARNT PAS-B domain, which has the same
201 tertiary structure as HIF-2α PAS-B but contains smaller internal cavities which total 150
202 $Å^3$ in volume. We have previously used NMR-based fragment screening, isothermal
203 titration calorimetry (ITC) and microscale thermophoresis (MST) to identify several
204 compounds which bind ARNT PAS-B with micromolar affinities; two of these, **KG-548**
205 and **KG-655**, are further known to disrupt ARNT PAS-B interactions with coactivator
206 proteins (3). However, the binding modes of these compounds remain unclear without
207 co-complex structures solved by NMR or X-ray crystallography. By acquiring pressure
208 NMR data, we confirmed that ARNT PAS-B retains a flexibly-accessible interior cavity
209 with a stdev(c_i [^{15}N]) of 84 x ($2x10^{-10}$ ppm/bar^2) (**Figs. 4B and S5**). Repeating these in
210 the presence of the **KG-655** ligand, we observed a drop of stdev(c_i [^{15}N]) to 51 x ($2x10^{-10}$
211 ppm/bar^2), consistent with an interior binding location for this small ligand. However,
212 similar studies with the larger **KG-548** ligand show a stdev(c_i [^{15}N]) comparable to the
213 apo protein, with an observed value of 96 x ($2x10^{-10}$ ppm/bar^2) that suggest that this
214 compound binds in a different mode than **KG-655**, potentially outside the cavity.
215
216 *Use of pressure NMR to reveal ligand binding – Poorly-defined cases:* As a next
217 demonstration of this approach, we examined its utility for PAS domains outside the
218 hypoxia response and with less well understood regulation. We considered two such
219 targets, the first of which is the N-terminal PAS domain of human PAS kinase (PASK
220 PAS-A), a PAS domain-regulated serine/threonine kinase conserved among eukaryotes
221 (4, 41). This domain has a canonical PAS structure, with a five-stranded β-sheet
222 flanked to one side by several α helices (27, 42). While only a small surface groove of

223 73 Å^3 could be identified near the F/G loop in the representative member of the solution
224 structure ensemble, an NMR-based fragment screen of 750 compounds identified
225 several small molecules that bound with micromolar affinity to PASK PAS-A (4).
226 Chemical shift perturbations suggest that these compounds bind within the domain
227 interior, analogous to the subsequently-discovered HIF-2α and ARNT PAS-B binding
228 compounds, but the lack of a PASK PAS-A/ligand complex structure leaves this as an
229 open issue. To address this, we used high-pressure NMR to examine the response of
230 PASK PAS-A in its apo form and when saturated with two compounds with K_D=13-24
231 μM affinities (**KG-535**, **KG-571 =** compounds **1 and 2** in (4)). For the apo- protein, we
232 observed a stdev(c_i [^{15}N]) of 45 x (2x10^{-10} ppm/bar^2); from the correlation identified in
233 our test set (**Fig. 3**), this value predicts that PASK PAS-A contains a total void volume
234 slightly smaller than the 3345 Å^3 calculated from the representative member of the
235 solution structure ensemble (**Figs. 5A and S6**). Upon addition of the **KG-535** and **KG-**
236 **571** ligands, we observed smaller stdev(c_i [^{15}N]) values for both ligand-bound forms (35
237 and 28 x (2x10^{-10} ppm/bar^2) for the **KG-535** and **KG-571**-bound forms, respectively; **Fig.**
238 **5A**). These data suggest that both compounds bind within the void volume of PASK
239 PAS-A, with **KG-571** having a greater effect on the domain adopting alternative folded
240 conformations.

241 We completed our analyses of natural proteins by examining a prokaryotic PAS
242 domain from a novel histidine kinase from *Rhizobium etli (43)*, chosen because of a).
243 homology with the light-regulated PAS-histidine kinase EL346 (44, 45) and b). the
244 widespread usage of small molecule ligands to control the structure and function of
245 PAS-containing proteins (including many histidine kinases (28)). While no experimental

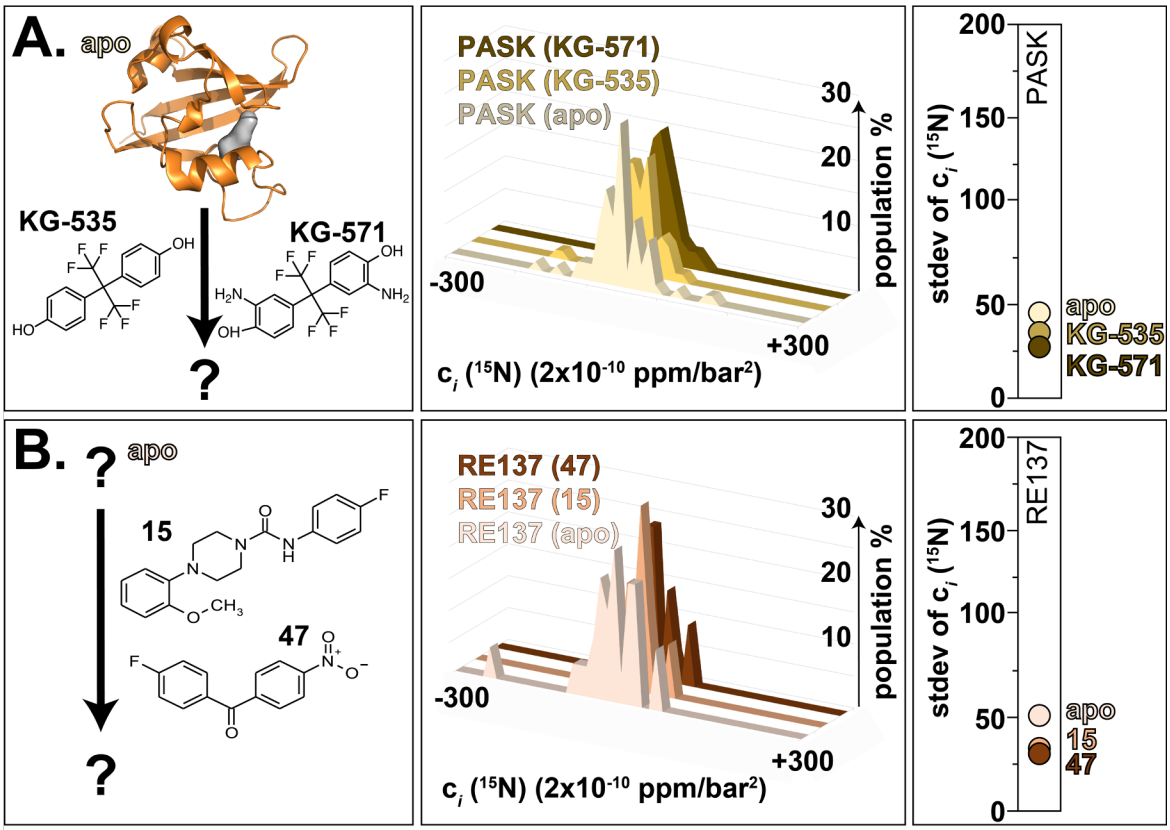

Fig. 5: High pressure NMR can provide ligand binding information, even in the absence of structural information. (Left panel) A. Three-dimensional structure of PASK PAS-A (PDB: 1ll8) and structures of two binding compounds, **KG-535** and **KG-571** (4). No high-resolution structures of protein/ligand complexes are available. B. No structural information is available for the RE137 PAS domain, in either apo and bound forms. Structures of two binding compounds identified by ^{19}F ligand-detected NMR screening, **15** and **47**. (Central panel) Histograms of the ^{15}N c_i coefficients measured on apo- and holo forms of PASK and RE137 PAS domains using the approach diagrammed in **Fig. 2**. (Right panel) Comparisons of the stdev(c_i [^{15}N]) values of PASK and RE137 apo- and ligand-bound samples.

structures are available of this domain, which we designate RE137, we hypothesized that the structure may contain a cavity that can accommodate a small molecule given precedence from other PAS domains (28). Pressure NMR analyses of the apo protein support this possibility, returning a stdev(c_i [^{15}N]) value of 53 x (2x10^{-10} ppm/bar^2) (**Figs. 5B and S7**), suggesting the presence of a 2800 Å^3 total void volume from the correlation seen in **Fig. 3**. To provide an initial assessment of small molecule binding,

252 we used an R_2-filtered ^{19}F NMR assay to search for binders of RE137 within a 100-
253 member library of fluorinated compounds (46). For this study, we used two of those
254 compounds **15** and **47**; pressure analyses of RE137 saturated with each compound
255 exhibited decreases of 43 and 39% with stdev(c_i [^{15}N]) values of 30 and 32 x ($2x10^{-10}$
256 ppm/bar^2) for the **15**- and **47**-bound forms, respectively. We interpret these data to
257 indicate that RE137 can bind both compounds in such a way as to reduce the total void
258 volume of the protein.
259
260 <u>*Use of pressure NMR to probe artificially-designed proteins:*</u> As a final demonstration of
261 the utility of this method, we examined its utility in investigating the flexibility of an
262 artificially-designed protein, CA01 (47). With advances in protein engineering enabling
263 the development of artificial ligand-binding biosensor proteins (48), we sought to probe
264 whether a protein such as CA01 – which contains a completely-enclosed 75 Å^3 cavity
265 within approximately 3500 Å^3 of total void volume – might show similar pressure-
266 dependent non-linear chemical shift changes as natively-evolved counterparts. From
267 our correlation in **Fig. 3** we expected to see a stdev(c_i [^{15}N]) value of approximately 75
268 ($2x10^{-10}$ ppm/bar^2) from a natural protein with void volume of this size, but we instead
269 observed a substantially smaller value of 38 x ($2x10^{-10}$ ppm/bar^2) (**Fig. 6**). As CA01 is
270 extremely stable to thermal and chemical denaturation, requiring 5 M guanidinium
271 hydrochloride to exhibit a complete thermal melt with a 75°C T_m (47), we view this
272 observation supports our prior findings with lysozyme suggesting thermostable proteins
273 being less able to adopt alternative conformations (**Fig. 3**).
274

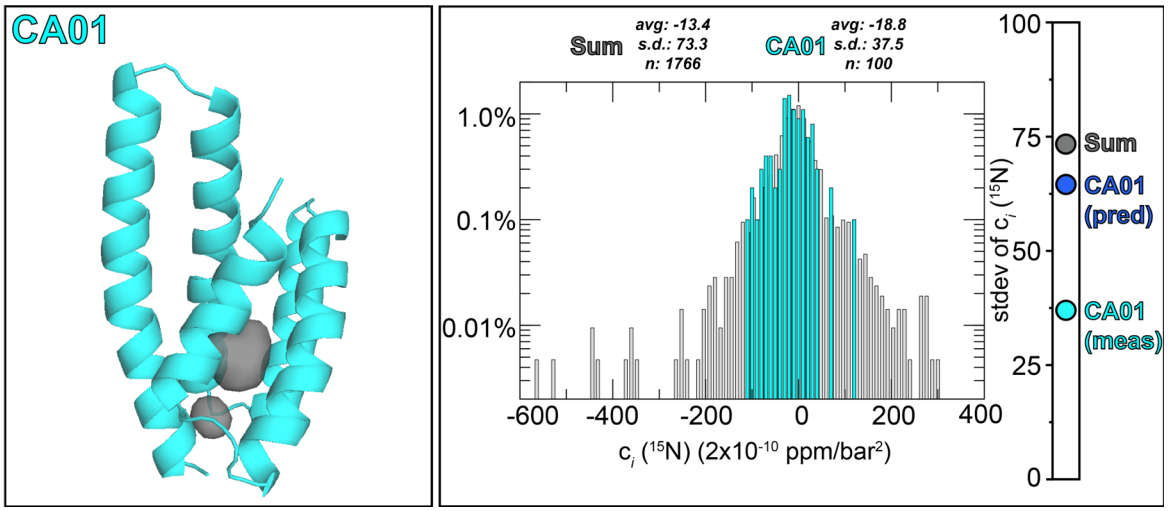

Fig. 6: High pressure NMR of a highly thermostable, cavity-containing designed protein gives smaller non-linear pressure dependent chemical shift changes than anticipated from void volume. (Left panel) Three-dimensional structure of CA01 (PDB: 5e6g), highlighting the location of a 75 Å^3 internal cavity within a total void volume of 3500 Å^3. (Right panel) Histogram of the ^{15}N c_i coefficients measured on CA01 (aqua) and 22 apo proteins listed in **Fig. S1** (grey) using the approach diagrammed in **Fig. 2**. Bar on right indicates stdev(c_i [^{15}N]) measured for CA01 (aqua) and 22 apo proteins (grey), along with the value predicted for CA01 from the correlation in **Fig 3.**

Discussion

In the present work, we successfully used non-linear pressure-induced chemical shift effects at backbone amide ^{1}H and ^{15}N nuclei to rapidly assess whether proteins contain substantial total void volumes within them and if these can bind small molecule ligands. A particular strength of our implementation is its simplicity, as 1). the necessary U-^{15}N labeled samples are routinely available from *E. coli* or *in vitro* expression systems, 2). the NMR data needed are easily obtained from conventional ^{1}H/^{15}N HSQC experiments recorded at variable pressures using commercially-available pressure sources, and 3). straightforward data analyses are done by simply tracking peak locations without requiring chemical shift assignments, eliminating one of the more time-consuming steps of NMR analyses.

286 Mechanistically, we attribute the non-linear chemical shift changes with
287 increasing pressure to site-specific conformational changes of the protein from the
288 native folded structural ensemble (N) to the low-lying excited state ensemble (N') with
289 smaller volume (17). Our specific choice of backbone amide ^{1}H and ^{15}N chemical shifts
290 give us complementary views of these transitions (**Fig. S3**), as ^{1}H chemical shifts are
291 dominated by hydrogen bonding to these atoms, while ^{15}N chemical shifts are
292 influenced by a broader set of factors, including changes in hydrogen bonding (to the
293 both the adjacent ^{1}H and carbonyl CO groups), local backbone dihedrals and sidechain
294 conformations (49). Of note, non-linear pressure-dependent chemical shift changes at
295 these sites are much smaller in peptides or intrinsically disordered proteins (50).
296 Indeed, non-linear pressure-dependent amide ^{1}H and ^{15}N chemical shift changes have
297 long been recognized as reflecting idiosyncratic differences among proteins (51). Prior
298 work by one of our groups (K.A.) suggested that this property depends on the density of
299 cavities within the protein (25); here we both expand the number of proteins examined
300 from that work and suggest that contributions likely arise from small and large voids
301 distributed throughout the protein which collectively give rise to the total void volumes
302 calculated by the program ProteinVolume (5) given the excellent correlation we present
303 in **Fig. 3**. We posit that such an approach reveals aspects that will be correlated to the
304 ability of proteins to be similarly affected by allosteric regulators and other ligands,
305 linking these non-linear chemical shift effects to the potential for small molecule binding
306 and regulation.
307 In addition to getting information on the total void volume within a given protein,
308 the strengths of this analysis include the ability to identify ligands which bind to a given

309 target and (de)stabilize one conformation in a way that may provide allosteric switching.
310 As such, this provides trivially-accessible mode-of-action information that is often
311 challenging to get otherwise without resources such as compounds of known binding
312 location for competition studies or cavity-filling point mutants. Our comparison of two
313 cavity-perturbing effects for HIF-2α PAS-B – small molecule binding and cavity-filling
314 redesign – produced opposite effects, with the small molecules stabilizing the native fold
315 and mutations destabilizing in our pressure NMR analyses. We underscore that these
316 effects occur despite minor changes in ground state structure among structures of the
317 apo, ligand-bound and redesigned proteins (1, 6, 40). As such, pressure-dependent
318 chemical shift effects give us insight into details of the packing and dynamics of proteins
319 more simply than most other methods can provide. Finally, our comparison of the
320 ARNT PAS-B binding compounds **KG-548** and **KG-655** provide an outstanding example
321 of two molecules known to bind by NMR chemical shift perturbations, but with limited
322 binding site information previously available. With high pressure NMR, we predict
323 different binding modes between the two ligands, with **KG-655** binding primarily to an
324 internal site while **KG-548** binds externally.

325 We emphasize that our approach requires very little preliminary information on
326 the target, needing neither structures nor site-specific chemical shift assignments, as
327 demonstrated by the hPASK PAS-A and RE137 cases. In both cases, the stdev(c_i [^{15}N])
328 values of the apo proteins exceed the value of approximately 50 (2×10^{-10} ppm/bar^2)
329 value we suggest as indicating potential ligand binding, and in both cases, we were able
330 to quickly identify compounds which bound to these proteins with micromolar affinities
331 from small libraries using either protein- or ligand-detected NMR methods. Our high

332 pressure NMR results quickly add important information, showing that these compounds
333 bind in such a way as to reduce the flexibility of the receptors, probably through binding
334 internal cavities.

335 Finally, while our demonstrations have focused on relatively small ~15 kDa ligand
336 binding domains, we note that several routes can expand the reach of this approach to
337 larger and more complex systems by reducing spectral complexity. Options include
338 biochemical approaches to generate proteins with simplified ^{15}N-labeling patterns, as
339 provided by amino-acid type selective or sortase/intein-based domain-specific
340 techniques, to ^{1}H/^{15}N TROSY or HNCO-type pulse sequences which can provide
341 improved resolution. Additional options may be provided by the improved NMR signal
342 properties of ^{1}H/^{13}C methyl signals for similar pressure-sensitive studies, particularly
343 given the placement of these residues near internal cavities.

344

345 **Materials and Methods**

346 Proteins were expressed in *E. coli* with uniform ^{15}N labeling, purified using
347 combination of Ni(II) affinity and gel filtration chromatography before being exchanged
348 into a barostatic Tris:phosphate buffer mix which limits pH changes during
349 pressurization (52) and concentrated to 100-700 μM for NMR spectroscopy. ^{1}H/^{15}N
350 HSQC spectra were acquired at increasing pressures from 20-2500 bar, interleaving
351 each high pressure spectrum with a low pressure (20 bar) spectrum to confirm that
352 changes in peak locations and intensities were reversible. All NMR data were
353 processed with NMRFx (One Moon Scientific) (53, 54) and analyzed with NMRViewJ
354 (One Moon Scientific) (53, 55). After individually processing each spectrum, we picked

peaks and tracked their changes in chemical shifts as a function of pressure. Separately handling movement in the ^{1}H and ^{15}N dimensions, we fit these trends to a second-order polynomial equation (**Eq. 1, Fig. 2**). Protein volume analyses of single isolated cavities utilized *cavfinder* (1), while measurements of total void volumes utilized *ProteinVolume* (5). Additional detailed procedures are found in **SI Materials and Methods**.

Acknowledgements

The authors gratefully acknowledge useful discussions and software contributions from Bruce Johnson (ASRC), assistance in sample preparation from Dong Lee and Andrew Palacios (ASRC), and comments on manuscript by Xingjian Xu (ASRC) and Hua Li (Chinese Academy of Sciences, Shanghai). U-^{15}N samples of FABP proteins were generously provided by May Poh Lai and Ruth Stark (CCNY); plasmids of Brd proteins were generously provided by Elena Rusinova and Ming-Ming Zhou (Mt. Sinai Medical Center) while plasmids of CA01 were generously shared by T.M. Jacobs and Brian Kuhlman (UNC Chapel Hill). This work was supported by NSF grant MCB-1818148 (K.H.G.) and NIH grant R01 GM106239 (K.H.G.), as well as a postdoctoral fellowship to D.G. (number B3X from the Fonds de Recherche Québec – Nature et Technologie (FRQNT)).

375 **References**

376 1. T. H. Scheuermann *et al.*, Allosteric inhibition of hypoxia inducible factor-2 with
377 small molecules. *Nat Chem Biol* **9**, 271-276 (2013).

378 2. T. H. Scheuermann *et al.*, Isoform-Selective and Stereoselective Inhibition of
379 Hypoxia Inducible Factor-2. *J Med Chem* **58**, 5930-5941 (2015).

380 3. Y. Guo *et al.*, Regulating the ARNT/TACC3 axis: multiple approaches to
381 manipulating protein/protein interactions with small molecules. *ACS Chem Biol* **8**,
382 626-635 (2013).

383 4. C. A. Amezcua, S. M. Harper, J. Rutter, K. H. Gardner, Structure and interactions
384 of PAS kinase N-terminal PAS domain: model for intramolecular kinase regulation.
385 *Structure* **10**, 1349-1361 (2002).

386 5. C. R. Chen, G. I. Makhatadze, ProteinVolume: calculating molecular van der
387 Waals and void volumes in proteins. *BMC Bioinformatics* **16**, 101 (2015).

388 6. T. H. Scheuermann *et al.*, Artificial ligand binding within the HIF2alpha PAS-B
389 domain of the HIF2 transcription factor. *Proc Natl Acad Sci U S A* **106**, 450-455
390 (2009).

391 7. J. Ashkani, D. J. Rees, The Critical Role Of VP1 In Forming The Necessary
392 Cavities For Receptor-mediated Entry Of FMDV To The Host Cell. *Sci Rep* **6**,
393 27140 (2016).

394 8. J. Mondal, N. Ahalawat, S. Pandit, L. E. Kay, P. Vallurupalli, Atomic resolution
395 mechanism of ligand binding to a solvent inaccessible cavity in T4 lysozyme. *PLoS*
396 *Comput Biol* **14**, e1006180 (2018).

397 9. A. S. Tanwar, V. D. Goyal, D. Choudhary, S. Panjikar, R. Anand, Importance of
398 Hydrophobic Cavities in Allosteric Regulation of Formylglycinamide Synthetase:
399 Insight from Xenon Trapping and Statistical Coupling Analysis. *PLoS ONE* **8**,
400 e77781 (2013).

401 10. C. D. Andersson, B. Y. Chen, A. Linusson, Mapping of ligand-binding cavities in
402 proteins. *Proteins: Structure, Function, and Bioinformatics* **78**, 1408-1422 (2010).

403 11. Z. Guo *et al.*, Identification of Protein–Ligand Binding Sites by the Level-Set
404 Variational Implicit-Solvent Approach. *Journal of Chemical Theory and*
405 *Computation* **11**, 753-765 (2015).

406 12. Z. Zhang, Y. Li, B. Lin, M. Schroeder, B. Huang, Identification of cavities on protein
407 surface using multiple computational approaches for drug binding site prediction.
408 *Bioinformatics* **27**, 2083-2088 (2011).

409 13. J. Dundas *et al.*, CASTp: computed atlas of surface topography of proteins with
410 structural and topographical mapping of functionally annotated residues. *Nucleic*
411 *Acids Res* **34**, W116-W118 (2006).

412 14. J.-K. Kim *et al.*, BetaVoid: Molecular voids via beta-complexes and Voronoi
413 diagrams: BetaVoid: Molecular Voids. *Proteins: Structure, Function, and*
414 *Bioinformatics* **82**, 1829-1849 (2014).

15. K. P. Tan, T. B. Nguyen, S. Patel, R. Varadarajan, M. S. Madhusudhan, Depth: a web server to compute depth, cavity sizes, detect potential small-molecule ligand-binding cavities and predict the pKa of ionizable residues in proteins. *Nucleic Acids Res* **41**, W314-W321 (2013).

16. Y. Xu *et al.*, CavityPlus: a web server for protein cavity detection with pharmacophore modelling, allosteric site identification and covalent ligand binding ability prediction. *Nucleic Acids Res* **46**, W374-W379 (2018).

17. D. A. Keedy *et al.*, An expanded allosteric network in PTP1B by multitemperature crystallography, fragment screening, and covalent tethering. *Elife* **7**, e36307 (2018).

18. C. Mattos, D. Ringe, Locating and characterizing binding sites on proteins. *Nat Biotechnol* **14**, 595-599 (1996).

19. M. Gerstein, C. Chothia, Packing at the protein-water interface. *Proc Natl Acad Sci U S A* **93**, 10167-10172 (1996).

20. K. Akasaka, Exploring the entire conformational space of proteins by high-pressure NMR. *Pure and Applied Chemistry* **75**, 927-936 (2003).

21. K. Akasaka, Probing conformational fluctuation of proteins by pressure perturbation. *Chem Rev* **106**, 1814-1835 (2006).

22. J. A. Caro, A. J. Wand, Practical aspects of high-pressure NMR spectroscopy and its applications in protein biophysics and structural biology. *Methods* **148**, 67-80 (2018).

23. J. Roche *et al.*, Cavities determine the pressure unfolding of proteins. *Proc Natl Acad Sci U S A* **109**, 6945-6950 (2012).

24. H. R. Kalbitzer *et al.*, Intrinsic allosteric inhibition of signaling proteins by targeting rare interaction states detected by high-pressure NMR spectroscopy. *Angew Chem Int Ed Engl* **52**, 14242-14246 (2013).

25. K. Akasaka, H. Li, Low-Lying Excited States of Proteins Revealed from Nonlinear Pressure Shifts in ^{1}H and ^{15}N NMR. *Biochemistry* **40**, 8665-8671 (2001).

26. M. Gross, R. Jaenicke, Proteins under pressure. The influence of high hydrostatic pressure on structure, function and assembly of proteins and protein complexes. *Eur J Biochem* **221**, 617-630 (1994).

27. B. L. Taylor, I. B. Zhulin, PAS domains: internal sensors of oxygen, redox potential, and light. *Microbiol Mol Biol Rev* **63**, 479-506 (1999).

28. J. T. Henry, S. Crosson, Ligand-binding PAS domains in a genomic, cellular, and structural context. *Annu Rev Microbiol* **65**, 261-286 (2011).

29. A. Losi, K. H. Gardner, A. Moglich, Blue-Light Receptors for Optogenetics. *Chem Rev* **118**, 10659-10709 (2018).

30. K. Akasaka *et al.*, Pressure response of protein backbone structure. Pressure-induced amide ^{15}N chemical shifts in BPTI. *Protein Sci* **8**, 1946-1953 (1999).

31. T. Knubovets, J. J. Osterhout, P. J. Connolly, A. M. Klibanov, Structure, thermostability, and conformational flexibility of hen egg-white lysozyme dissolved in glycerol. *Proc Natl Acad Sci U S A* **96**, 1262-1267 (1999).

32. Y. O. Kamatari, L. J. Smith, C. M. Dobson, K. Akasaka, Cavity hydration as a gateway to unfolding: an NMR study of hen lysozyme at high pressure and low temperature. *Biophys Chem* **156**, 24-30 (2011).

33. W. G. Kaelin, Jr., P. J. Ratcliffe, Oxygen sensing by metazoans: the central role of the HIF hydroxylase pathway. *Mol Cell* **30**, 393-402 (2008).

34. D. Wu, N. Potluri, J. Lu, Y. Kim, F. Rastinejad, Structural integration in hypoxia-inducible factors. *Nature* **524**, 303-308 (2015).

35. D. Wu *et al.*, Bidirectional modulation of HIF-2 activity through chemical ligands. *Nat Chem Biol* **15**, 367-376 (2019).

36. W. Chen *et al.*, Targeting renal cell carcinoma with a HIF-2 antagonist. *Nature* **539**, 112-117 (2016).

37. H. Cho *et al.*, On-target efficacy of a HIF-2alpha antagonist in preclinical kidney cancer models. *Nature* **539**, 107-111 (2016).

38. E. M. Wallace *et al.*, A Small-Molecule Antagonist of HIF2alpha Is Efficacious in Preclinical Models of Renal Cell Carcinoma. *Cancer Res* **76**, 5491-5500 (2016).

39. J. Key, T. H. Scheuermann, P. C. Anderson, V. Daggett, K. H. Gardner, Principles of ligand binding within a completely buried cavity in HIF2alpha PAS-B. *J Am Chem Soc* **131**, 17647-17654 (2009).

40. F. Corrêa, J. Key, B. Kuhlman, Kevin H. Gardner, Computational Repacking of HIF-2α Cavity Replaces Water-Based Stabilized Core. *Structure* **24**, 1918-1927 (2016).

41. J. Rutter, C. H. Michnoff, S. M. Harper, K. H. Gardner, S. L. McKnight, PAS kinase: An evolutionarily conserved PAS domain-regulated serine/threonine kinase. *Proceedings of the National Academy of Sciences* **98**, 8991-8996 (2001).

42. A. Möglich, R. A. Ayers, K. Moffat, Structure and Signaling Mechanism of Per-ARNT-Sim Domains. *Structure* **17**, 1282-1294 (2009).

43. V. Gonzalez *et al.*, The partitioned Rhizobium etli genome: Genetic and metabolic redundancy in seven interacting replicons. *Proceedings of the National Academy of Sciences* **103**, 3834-3839 (2006).

44. F. Correa, W. H. Ko, V. Ocasio, R. A. Bogomolni, K. H. Gardner, Blue light regulated two-component systems: enzymatic and functional analyses of light-oxygen-voltage (LOV)-histidine kinases and downstream response regulators. *Biochemistry* **52**, 4656-4666 (2013).

45. G. Rivera-Cancel, W. H. Ko, D. R. Tomchick, F. Correa, K. H. Gardner, Full-length structure of a monomeric histidine kinase reveals basis for sensory regulation. *Proc Natl Acad Sci U S A* **111**, 17839-17844 (2014).

46. A. Vulpetti, U. Hommel, G. Landrum, R. Lewis, C. Dalvit, Design and NMR-Based Screening of LEF, a Library of Chemical Fragments with Different Local Environment of Fluorine. *J Am Chem Soc* **131**, 12949-12959 (2009).

47. T. M. Jacobs *et al.*, Design of structurally distinct proteins using strategies inspired by evolution. *Science* **352**, 687-690 (2016).

48. J. Feng *et al.*, A general strategy to construct small molecule biosensors in eukaryotes. *Elife* **4**, e10606 (2015).

49. D. Sitkoff, D. A. Case, Theories of chemical shift anisotropies in proteins and nucleic acids. *Progress in Nuclear Magnetic Resonance Spectroscopy* **32**, 165-190 (1998).

50. M. Beck Erlach *et al.*, Pressure dependence of side chain 13C chemical shifts in model peptides Ac-Gly-Gly-Xxx-Ala-NH2. *J Biomol NMR* **69**, 53-67 (2017).

51. R. Kitahara, K. Hata, H. Li, M. P. Williamson, K. Akasaka, Pressure-induced chemical shifts as probes for conformational fluctuations in proteins. *Progress in Nuclear Magnetic Resonance Spectroscopy* **71**, 35-58 (2013).

52. R. J. Quinlan, G. D. Reinhart, Baroresistant buffer mixtures for biochemical analyses. *Anal Biochem* **341**, 69-76 (2005).

53. B. A. Johnson, From Raw Data to Protein Backbone Chemical Shifts Using NMRFx Processing and NMRViewJ Analysis. *Methods Mol Biol* **1688**, 257-310 (2018).

54. M. Norris, B. Fetler, J. Marchant, B. A. Johnson, NMRFx Processor: a cross-platform NMR data processing program. *J Biomol NMR* **65**, 205-216 (2016).

55. B. A. Johnson, R. A. Blevins, NMRView: a computer program for the visualization and analysis of NMR data. *J Biomol NMR* **4**, 603-614 (1994).