ChemPubSoc Europe

DOI: 10.1002/cbic.201402585

CHEMBIOCHEM Communications

Conformation of Inhibitor-Free HIV-1 Protease Derived from NMR Spectroscopy in a Weakly Oriented Solution

Julien Roche,* John M. Louis,* and Ad Bax*[a]

Flexibility of the glycine-rich flaps is known to be essential for catalytic activity of the HIV-1 protease, but their exact conformations at the different stages of the enzymatic pathway remain subject to much debate. Although hundreds of crystal structures of protease–inhibitor complexes have been solved, only about a dozen inhibitor-free protease structures have been reported. These latter structures reveal a large diversity of flap conformations, ranging from closed to semi-open to wide open. To evaluate the average structure in solution, we measured residual dipolar couplings (RDCs) and compared these to values calculated for crystal structures representative of the closed, semi-open, and wide-open states. The RDC data clearly indicate that the inhibitor-free protease, on average, adopts a closed conformation in solution that is very similar to the inhibitor-bound state. By contrast, a highly drug-resistant protease mutant, PR20, adopts the wide-open flap conformation.

The human immunodeficiency virus type 1 (HIV-1) protease (PR) is an aspartic hydrolase that functions as an obligatory homodimer with 99 amino acids in each subunit. Each monomer contains a glycine-rich flap segment (residues 44–57) that folds into an antiparallel β-sheet and covers the active site (Figure 1 A). The flexibility of the flaps is believed to control the access of substrates and inhibitors to the active site and is therefore recognized as essential for enzymatic activity.[1,2] Despite intense research over the past 20 years, the detailed role of the flaps in the substrate binding mechanism remains subject to much debate. Whereas most of the current mechanistic models propose that wide opening of the flaps gives access to the active site,[3–5] a recent molecular dynamics (MD) simulation suggests a translational procession of the substrate into the active site while maintaining the flaps in closed or semi-open conformations.[6] The access pathway of substrate to the active site of HIV-1 PR is of fundamental interest for gaining an improved understanding of how this class of enzymes channels its substrates and is also particularly important for the design of high-affinity future-generation inhibitors.

Several hundred crystal structures of protease–inhibitor complexes have been deposited in the Protein Data Bank (PDB), and in all of these complexes, the flaps adopt closed conformations. On the other hand, far fewer structural data are available in the absence of inhibitor, in part due to the problems associated with autoproteolysis and decreased stability of PR in the absence of inhibitors. These inhibitor-free structures show a variety of flap conformations, ranging from closed to semi-open to wide open, in which the flap hairpin moieties have moved apart by several angstroms compared to the closed conformation.[7] Interestingly, differences between the different subtypes of HIV have also been reported. For example, although the unbound PR, isolated from subtype A HIV, has only been crystallized in a closed conformation, for subtype B, six inhibitor-free structures were found in a semi-open conformation, two in a wide-open conformation, and only a single one in the closed conformation.[8] These data cannot unambiguously distinguish whether this heterogeneity reflects the preferred flap conformations or results from crystal packing effects.

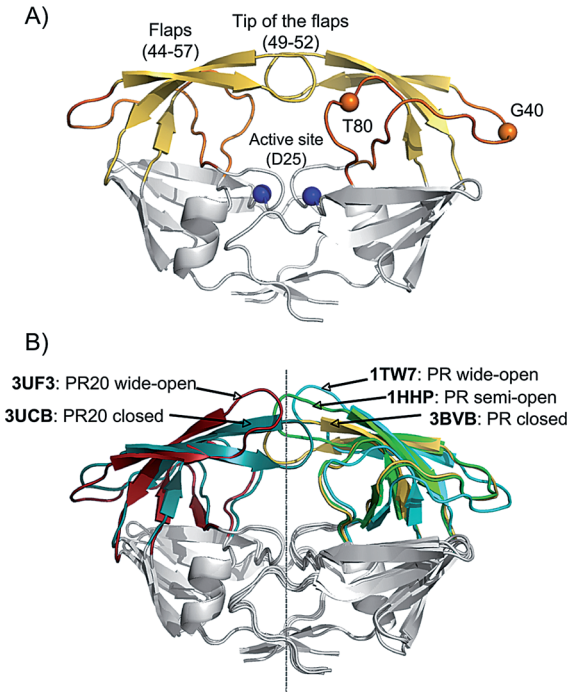

Figure 1. X-ray structures of HIV-1 PR. A) Structure of PR with flaps in the closed conformation (PDB ID: 3BVB), highlighting the flaps (yellow), the hinge regions (orange) and the active site (blue). B) Comparison of the different structural models used in this study, showing the flap conformations of PR in closed (3BVB, yellow), semi-open (1HHP, green) and wide-open (1TW7, cyan) conformations, as well as the PR20 structures used as reference models for the closed (3UCB, dark cyan) and wide-open (3UF3, red) flap conformations. All models have been superimposed to yield a minimal C^{α} rmsd for the core region (comprising residues 10–23, 62–73, and 87–93 of both monomers).

[a] *Dr. J. Roche, Dr. J. M. Louis, Dr. A. Bax*
Laboratory of Chemical Physics
National Institute of Diabetes and Digestive and Kidney Diseases
National Institutes of Health
Bethesda, MD 20892 (USA)
E-mail: julien.roche@nih.gov
johnl@niddk.nih.gov
bax@nih.gov

The present study addresses the question concerning the preferred flap orientations by using solution NMR spectroscopy to record residual dipolar couplings (RDCs) in a B-type active site mutant PR in the absence of inhibitor. These RDCs can be measured by imposing a very slight deviation from the random, isotropic distribution of macromolecules in an NMR sample and are very sensitive reporters on the time-averaged orientation of the corresponding internuclear vectors.[9,10] RDCs for the backbone N–H amide vectors were measured here for an inhibitor-free PR and a symmetric inhibitor DMP323-bound complex,[11] as well as for an unbound drug-resistant PR bearing 20 mutations (PR20) plus a D25N mutation at the active site.[12] Weak alignment of the NMR samples was obtained by the addition of squalamine, an antibiotic isolated from the dogfish shark, which was recently shown to form a lyotropic liquid crystal at very small volume fractions in water.[13] The lack of any direct interaction between squalamine and PR was evidenced by the complete absence of any detectable change in chemical shifts in the ^{1}H–^{15}N chemical shift correlation spectrum, which is an exquisitely sensitive reporter of such interactions when present. Although amide RDCs alone are typically insufficient to build a protein structure de novo, they are exceptionally well suited to evaluate agreement between the actual state of the protein present in solution and the coordinates seen in different crystal structures; this is the approach taken in our study.

Here, we chose three crystal structures as reference models for representing the three different conformations of the flaps: PDB ID: 3BVB[14] for the closed conformation, 1HHP[15] as the most widely used model for the semi-open conformation, and 1TW7[16] for the wide-open conformation (Figure 1 B). To evaluate movement of the flaps, we defined the PR core as the region of the protein dimer for which predicted amide N–H vector orientations, and thereby the predicted $^{1}D_{NH}$ RDCs, are most invariant between these three structural models. For this invariant core region, excellent agreement between measured $^{1}D_{NH}$ RDCs and best-fitted values was obtained for all three X-ray structures and the experimental data collected for unbound and DMP323-bound PR, as well as for the inhibitor-free PR20 mutant, as reflected in *Q*-factors close to 20% (reported as *Q*-Core, Figure 2). This core region comprises the first and last two β-sheets (residues 10–23 and 62–73), as well as the sole α-helix (residues 87–93). The five alignment tensor parameters determined for each structural model from a singular value decomposition (SVD) fit to the core region were then fixed and used to predict $^{1}D_{NH}$ RDCs for the flap region (Table 1). Here, we defined the flaps as extended regions that included both the Thr80 loop and Gly40 loop (Figure 1 A), as these two hinge segments are also involved in the opening and closing of the flaps. They are relatively invariant between the different inhibitor-free and inhibitor-bound X-ray structures (C^{α} root mean square difference average of <0.4 Å).

As expected for an inhibitor-bound complex, the flap RDCs measured for the PR in the presence of DMP323 agreed much better with the coordinates of the closed X-ray structure, 3BVB, than those of the semi-open (1HHP) and wide-open (1TW7) reference structures, yielding *Q*-factors of 18.6, 32.3, and 75.8%, respectively (Figure 2). Interestingly, much better agreement was also obtained when comparing the RDCs of the inhibitor-free PR flaps to the closed structure, 3BVB (*Q*-flap =

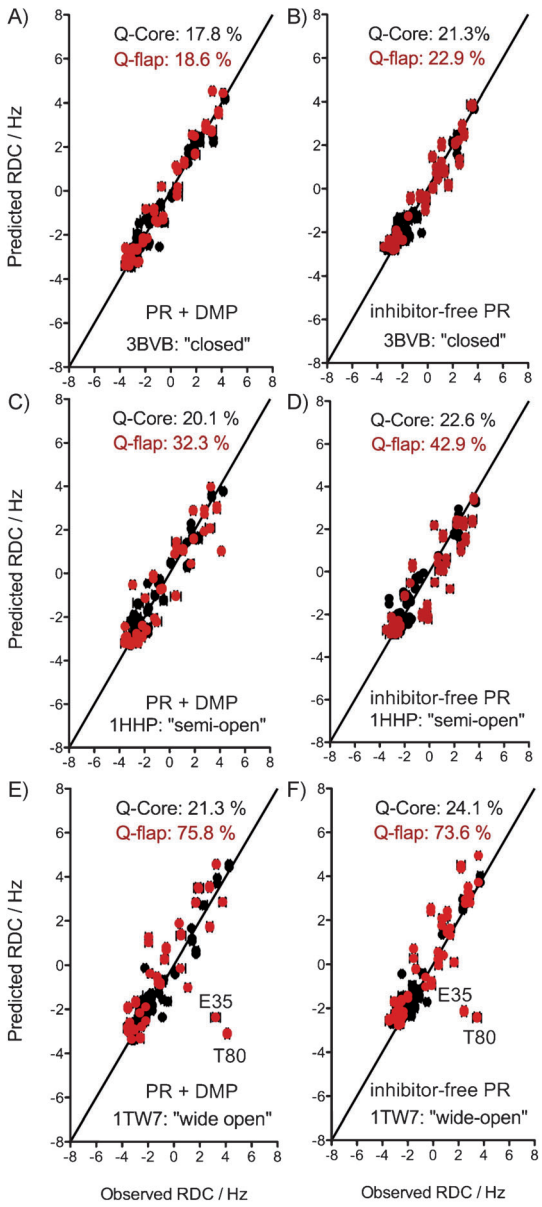

Figure 2. Fit of the NH RDCs measured for the inhibitor-free and DMP323-bound PR to the closed (A and B), semi-open (C and D) and wide-open (E and F) structural models. The fit of the core region RDCs (residues 10–23, 62–73, and 87–93) are shown in black. RDCs corresponding to the flaps (residues 30–61 and 74–84) are displayed in red, with predicted values derived by using the alignment tensor parameters of the core region. The quality factor, *Q*, reflects the rmsd between observed RDCs and those predicted by the structure, normalized for the range of RDCs observed for each sample: $Q = \mathrm{rms}(D_{calcd} - D_{obs})/\mathrm{rms}(D_{obs})$, where rms is the root-mean-square function, and D_{obs} and D_{calcd} refer to the observed RDC value and the RDC calculated based on the atomic coordinates, when using a best-fitted alignment tensor.[17] To avoid an effect from the non-random distribution of the ^{15}N–^{1}H vectors in the protein, $\mathrm{rms}(D_{obs})$ was actually calculated as $\{D_a^2[4+3R^2]/5\}^{1/2}$, where *Da* and *R* are the magnitude and rhombicity of the best-fitted alignment tensor.[18] The experimental uncertainty in D_{obs} was very small (±0.15 Hz) but, due to the uncertainty in the exact positions of the hydrogens, used for deriving D_{calcd}; even for a perfect X-ray structure, the lower limit on *Q* is ~0.15.

Table 1. SVD-derived alignment tensor parameters obtained by a fit of the $^1D_{NH}$ RDCs to the core region of the corresponding structural models.[a]

	Da [Hz]	Rh	Euler (*x*)	Euler (*y*)	Euler (*z*)
PR + DMP[b] 3BVB	2.92	0.13	43.8	−20.8	53.6
PR + DMP[b] 1HHP	2.91	0.12	40.2	−17.5	56.3
PR + DMP[b] 1TW7	2.84	0.15	43.3	−21.1	54.9
PR apo[c] 3BVB	2.91	0.11	41.9	−19.1	52.8
PR apo[c] 1HHP	2.98	0.11	39.0	−14.2	53.8
PR apo[c] 1TW7	2.79	0.12	41.2	−20.0	51.5
PR20 apo[d] 3UF3	3.12	0.17	40.5	−18.8	51.6
PR20 apo[d] 3UCB	3.01	0.16	39.6	−14.4	46.3

[a] The core region comprises the first and last two β-sheets (residues 10–23 and 62–73), as well as the C-terminal helix (residues 87–93). The magnitude of the alignment tensor obtained for the apo PR and PR20 data sets was scaled up according to the ^{2}H quadrupole splitting measured for the PR + DMP323 sample (25.6 Hz, compared to 22 and 19 Hz for the apo PR and PR20, respectively). [b] RDC data collected for PR bound to DMP323. [c] RDC data collected for inhibitor-free PR. [d] RDC data collected for inhibitor-free PR20.

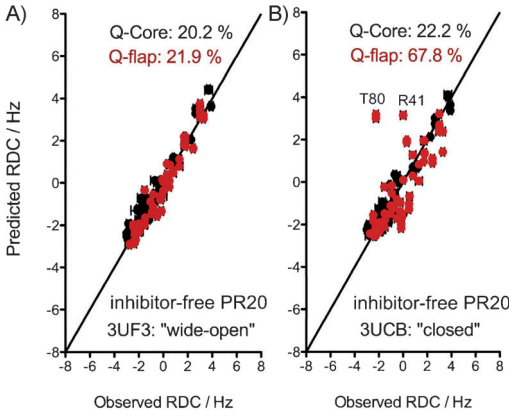

Figure 3. Correlation between $^1D_{NH}$ RDCs measured for the unbound PR20 and (A) wide-open or (B) closed structural models. RDCs corresponding to the core region are shown in black and were used to determine the alignment tensor through SVD fitting; flap RDCs are shown in red, with predicted values derived by using alignment tensor parameters obtained for the core residues.

22.9%) than for the semi-open (*Q*-flap = 42.9%) and wide-open (*Q*-flap = 73.6%) references (Figure 2). For both the inhibitor-free and DMP323-bound states, the hinge residues Thr80 and Glu35 displayed the largest deviations between the experimental and predicted values when the flap RDCs were compared to the wide-open structure, 1TW7 (Figure 2E and F). In this respect, it is important to note that, even for the very best X-ray structures of any protein, RDC *Q*-factors will rarely fall below about 15–20%, mostly limited by the fact that X-ray maps do not contain adequate electron density for defining the precise position of hydrogens, thus, these are typically added by model-building, assuming idealized geometry.[13,18]

To verify that the backbone RDCs could capture a truly open flap conformation, we also measured RDCs for a protease bearing 20 mutations (PR20), found clinically in a patient with high protease drug resistance.[19] PR20 exhibited a threefold higher dimer dissociation constant, a similar catalytic constant (k_{cat}), an ~13-fold higher K_m value for the hydrolysis of a synthetic substrate, and more than three orders of magnitude decreased affinity for the clinical protease inhibitors darunavir and saquinavir, relative to the wild-type PR.[12] The crystal structure of inhibitor-free PR20 reveals wide-open conformations of the flaps, similar to those seen in 1TW7, whereas in the inhibitor-bound PR20, the flaps were found in their usual, closed conformation.[20]

We measured the $^1D_{NH}$ RDCs for unbound PR20 under the same conditions as for inhibitor-free and DMP323-bound PR and used the two corresponding PR20 crystal structures, 3UCB and 3UF3, as reference structures for the closed and wide-open conformations, respectively. As for PR, the alignment tensor parameters were determined by an SVD fit of the PR20 RDCs to the core region, and these parameters were subsequently used to predict the RDCs for the extended flap regions (Table 1). The experimental RDCs show much closer agreement with values predicted for the wide-open PR20 conformation (Figure 3A; *Q*-flap = 21.9%) than those for the closed conformation (Figure 3B; *Q*-flap = 67.8%). Therefore, in contrast to inhibitor-free PR, the RDCs measured for PR20 clearly indicated that the flaps adopt a wide-open conformation in the unbound state. The large difference between the inhibitor-free conformation of PR and PR20 likely arises from the mutations in the flap hinge region (E35D, M36I, and S37N), which eliminate several native interactions that have been postulated to impact the orientation of the flaps.[20]

NMR relaxation data measured for inhibitor-free PR revealed dynamic fluctuations of the flap tips on the μs–ms time scale,[21] which have been commonly interpreted as reflecting a dynamic equilibrium between semi-open and wide-open conformations.[3–5] Combining the crystallographic and NMR data with MD simulations, a popular model of the PR binding mechanism has emerged over the past years. In this model, the inhibitor-free protease adopts a semi-open conformation with rare and transient wide opening of the flaps that would allow smooth access of the natural polyproteins into the active site. After the substrate binds to the active site, the flaps close and then adopt the conformation observed in the protease–inhibitor complexes. However, it is important to realize that the NMR relaxation data cannot distinguish a semi-open to wide-open conformational transition from any other mode of structural fluctuation. The backbone RDCs measured in the present study clearly show that the average conformation of the inhibitor-free PR in solution is very similar to the closed conformation observed for PR–inhibitor complexes. The NMR relaxation data reported by Freedberg et al.[21] on the inhibitor-free PR can therefore be reinterpreted in view of the present data as reflecting transient switching between closed and a small population of semi-open, or possibly wide-open, conformations, rather than between semi-open and wide-open conformations. Indeed, a recent MD simulation study shows a transversal procession of the N-terminal cleavage site sequence (TFR/PR) through the active site without wide opening of the flaps.[6] A complete opening of the flaps, therefore, might not be required for the catalytic mechanism of PR. In this regard, it is

also interesting to note that the flap RDCs measured in the inhibitor-free state are not systematically reduced relative to those seen in the Core region, indicating that dynamic averaging of these RDCs, which is integrated over the time scale of milliseconds in such RDC measurements,[22,23] is minimal. Our data, therefore, paint a picture in which the flaps, with the exception of their tips, have a structural rigidity that is comparable to the core region of the dimer and predominantly adopt the closed conformation, even in the absence of substrate or ligands. RDCs predicted for the semi-open state differ only modestly from those of the closed state, and our RDC data suggest an upper limit of at most ~25% for the semi-open state. If the alternate state is the wide-open configuration, for which the predicted RDCs differ substantially more from the closed state, an upper limit of about 10% for the population of this state applies.

Our results demonstrate that measurement of RDCs provides a convenient method for identifying the loop conformations actually present in solution from a range of possibilities suggested by available crystal structures. This type of study is analogous to earlier work on other flexible protein systems, such as hemoglobin[24] and lysozyme.[25] For each of these, two different liquid crystalline media were used to prove that the state of the protein is not impacted by the alignment medium. For PR, it was difficult to find liquid crystalline media compatible with both the free and inhibitor-bound states of the protein, and we were only able to get high quality data in the newly discovered squalamine liquid crystal. However, the fact that we can observe the wide-open state for inhibitor-free PR20, versus the closed state for inhibitor-free PR, indicates that the protein is not forced to adopt any particular state by the presence of the liquid crystal.

The use of paramagnetic NMR, pseudo-contact shifts in particular, provides an alternate method for probing the structure of proteins subject to dynamic rearrangement,[26] such as the state of the HIV-1 protease flaps in solution. This method was recently used to investigate the structure of the flap covering the active site in the heterodimeric dengue virus protease NS2B-NS3, which was surprisingly also found to adopt the closed conformation in the inhibitor-free state.[27] This approach requires that paramagnetic lanthanide tags be linked at suitable positions to the protein but obviates the need for finding a suitable orienting medium needed for measurement of RDCs. In favorable cases, introduction of the paramagnetic tag also yields weak protein alignment without requiring an orienting medium, thus enabling unambiguous evaluation of the dynamic properties of the protein.[28] Both RDC and pseudo-contact shift methods, therefore, are particularly useful for probing the state of dynamic regions in isotropic solution, under conditions that can closely mimic the relevant physiological environment.

Experimental Section

Uniformly $^{2}H/^{13}C/^{15}N$-enriched (>98%) protease samples were purified from inclusion bodies by size-exclusion chromatography under denaturing conditions, followed by reversed-phase HPLC, as previously described.[29] Proteins were folded by dilution from a stock solution of 2 mg mL^{-1} in HCL (12 mM) in 6.6 volumes of acetate buffer (5 mM, pH 6.0), with or without DMP323 inhibitor, and then were dialyzed extensively in 20 mM sodium phosphate (pH 5.7) and concentrated. The protease constructs used in this study contained an additional active site D25N mutation to prevent autoproteolysis. The D25N mutation was introduced into PR20 by QuickChange mutagenesis.

The $^{1}D_{NH}$ RDCs were derived from the difference in $^{1}J_{NH}+{}^{1}D_{NH}$ splitting measured at 600 MHz using an ARTSY-HSQC experiment[30] on an isotropic sample and an aligned sample. All experiments were performed at 293 K. The alignment of the samples was obtained by the addition of 10 mg mL^{-1} squalamine and 5 mM hexan-1-ol, yielding a stable ^{2}H quadrupole splitting of ~20 Hz. The average experimental error in the measured $^{1}D_{NH}$ RDCs was 0.15 Hz. Alignment tensor parameters are listed in Table 1.

Acknowledgements

We thank Annie Aniana, James L. Baber, and Jinfa Ying for technical assistance and Dennis A. Torchia for helpful discussions. We acknowledge support from the National Institute of Diabetes and Digestive and Kidney Diseases Advanced Mass Spectrometry Core. This work was funded by the National Institutes of Health (NIH) Intramural Research Programs of the National Institute of Diabetes and Digestive and Kidney Diseases and by the Intramural AIDS-Targeted Antiviral Program of the Office of the Director, NIH.

Keywords: active sites · flap dynamics · HIV protease · liquid crystal NMR spectroscopy · residual dipolar coupling

[1] J. M. Louis, R. Ishima, D. A. Torchia, I. T. Weber in *Advances in Pharmacology, Vol. 55: HIV-1: Molecular Biology and Pathogenesis: Viral Mechanisms*, 2nd ed. (Ed.: K.-T. Jeang), Academic Press, San Diego, **2007**, p. 261.
[2] V. Y. Torbeev, H. Raghuraman, D. Hamelberg, M. Tonelli, W. M. Westler, E. Perozo, S. B. H. Kent, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 20982.
[3] V. Hornak, A. Okur, R. C. Rizzo, C. Simmerling, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 915.
[4] A. L. Perryman, J. H. Lin, J. A. McCammon, *Protein Sci.* **2004**, *13*, 1434.
[5] V. Hornak, C. Simmerling, *Drug Discovery Today* **2007**, *12*, 132.
[6] S. Kashif Sadiq, F. Noe, G. De Fabritiis, *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 20449.
[7] H. Heaslet, R. Rosenfeld, M. Giffin, Y.-C. Lin, K. Tam, B. E. Torbett, J. H. Elder, D. E. McRee, C. D. Stout, *Acta Crystallogr. Sect. D* **2007**, *63*, 866.
[8] A. H. Robbins, R. M. Coman, E. Bracho-Sanchez, M. A. Fernandez, C. T. Gilliland, M. Li, M. Agbandje-McKenna, A. Wlodawer, B. M. Dunn, R. McKenna, *Acta Crystallogr. Sect. D* **2010**, *66*, 233.
[9] J. H. Prestegard, C. M. Bougault, A. I. Kishore, *Chem. Rev.* **2004**, *104*, 3519.
[10] A. Bax, A. Grishaev, *Curr. Opin. Struct. Biol.* **2005**, *15*, 563.
[11] P. Y. S. Lam, P. K. Jadhav, C. J. Eyermann, C. N. Hodge, Y. Ru, L. T. Bacheler, J. L. Meek, M. J. Otto, M. M. Rayner, Y. N. Wong, C. H. Chang, P. C. Weber, D. A. Jackson, T. R. Sharpe, S. Ericksonviitanen, *Science* **1994**, *263*, 380.
[12] J. M. Louis, A. Aniana, I. T. Weber, J. M. Sayer, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 9072.
[13] A. S. Maltsev, A. Grishaev, J. Roche, M. Zasloff, A. Bax, *J. Am. Chem. Soc.* **2014**, *136*, 3752.
[14] J. M. Sayer, F. Liu, R. Ishima, I. T. Weber, J. M. Louis, *J. Biol. Chem.* **2008**, *283*, 13459.
[15] S. Spinelli, Q. Z. Liu, P. M. Alzari, P. H. Hirel, R. J. Poljak, *Biochimie* **1991**, *73*, 1391.

[16] P. Martin, J. F. Vickrey, G. Proteasa, Y. L. Jimenez, Z. Wawrzak, M. A. Winters, T. C. Merigan, L. C. Kovari, *Structure* **2005**, *13*, 1887.
[17] G. Cornilescu, J. L. Marquardt, M. Ottiger, A. Bax, *J. Am. Chem. Soc.* **1998**, *120*, 6836.
[18] A. Bax, *Protein Sci.* **2003**, *12*, 1.
[19] I. Dierynck, M. De Wit, E. Gustin, I. Keuleers, J. Vandersmissen, S. Hallenberger, K. Hertogs, *J. Virol.* **2007**, *81*, 13845.
[20] J. Agniswamy, C.-H. Shen, A. Aniana, J. M. Sayer, J. M. Louis, I. T. Weber, *Biochemistry* **2012**, *51*, 2819.
[21] D. I. Freedberg, R. Ishima, J. Jacob, Y.-X. Wang, I. Kustanovich, J. M. Louis, D. A. Torchia, *Protein Sci.* **2002**, *11*, 221.
[22] J. R. Tolman, *J. Am. Chem. Soc.* **2002**, *124*, 12020.
[23] O. F. Lange, N. A. Lakomek, C. Fares, G. F. Schroder, K. F. A. Walter, S. Becker, J. Meiler, H. Grubmüller, C. Griesinger, B. L. de Groot, *Science* **2008**, *320*, 1471.
[24] J. A. Lukin, G. Kontaxis, V. Simplaceanu, Y. Yuan, A. Bax, C. Ho, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 517.
[25] N. K. Goto, N. R. Skrynnikov, F. W. Dahlquist, L. E. Kay, *J. Mol. Biol.* **2001**, *308*, 745.
[26] L. Cerofolini, G. B. Fields, M. Fragai, C. F. G. C. Geraldes, C. Luchinat, G. Parigi, E. Ravera, D. I. Svergun, J. M. C. Teixeira, *J. Biol. Chem.* **2013**, *288*, 30659.
[27] L. de La Cruz, H. D. N. Thi, K. Ozawa, J. Shin, B. Graham, T. Huber, G. Otting, *J. Am. Chem. Soc.* **2011**, *133*, 19205.
[28] I. Bertini, C. Del Bianco, I. Gelis, N. Katsaros, C. Luchinat, G. Parigi, M. Peana, A. Provenzani, M. A. Zoroddu, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 6841.
[29] R. Ishima, D. A. Torchia, J. M. Louis, *J. Biol. Chem*. **2007**, *282*, 17190.
[30] N. C. Fitzkee, A. Bax, *J. Biomol. NMR* **2010**, *48*, 65.

Received: October 8, 2014
Published online on December 2, 2014