Angewandte Chemie
Eine Zeitschrift der Gesellschaft Deutscher Chemiker

Supporting Information

69451 Weinheim, Germany

Inferential NMR/X-ray-Based Structure Determination of a Dibenzo[*a*,*d*]cycloheptenone Inhibitor–p38α MAP Kinase Complex in Solution**

Valerie S. Honndorf, Nicolas Coudevylle, Stefan Laufer, Stefan Becker, Christian Griesinger, and Michael Habeck**

anie_201105241_sm_miscellaneous_information.pdf
anie_201105241_sm_supplementary_movie.avi

Supplementary figures

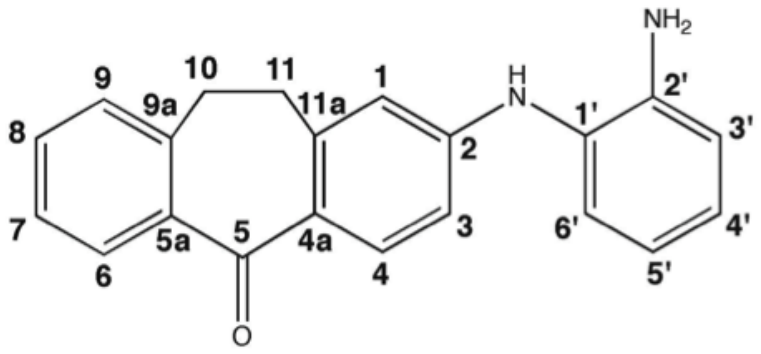


Figure S1: 10b inhibitor with numbering of the carbon atoms.

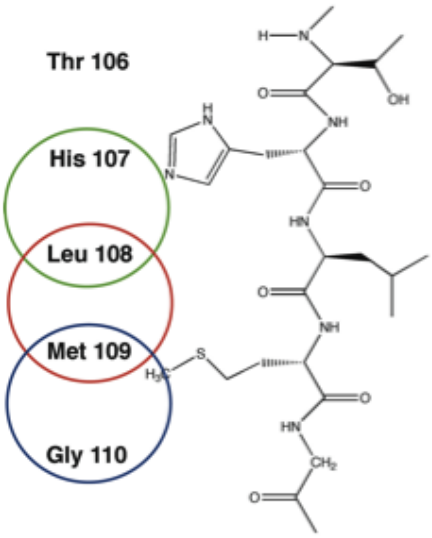


Figure S2: Sequence of the hinge region, labeling scheme of the protein samples is marked with circles, green circle the labeled amino acid pair ^{13}C,^{15}N His/^{15}N Leu, red circle $^{13}C^{15}N$ Leu/ ^{15}N Met labeled sample and blue circle $^{13}C^{15}N$ Met/^{15}N Gly labeled sample.

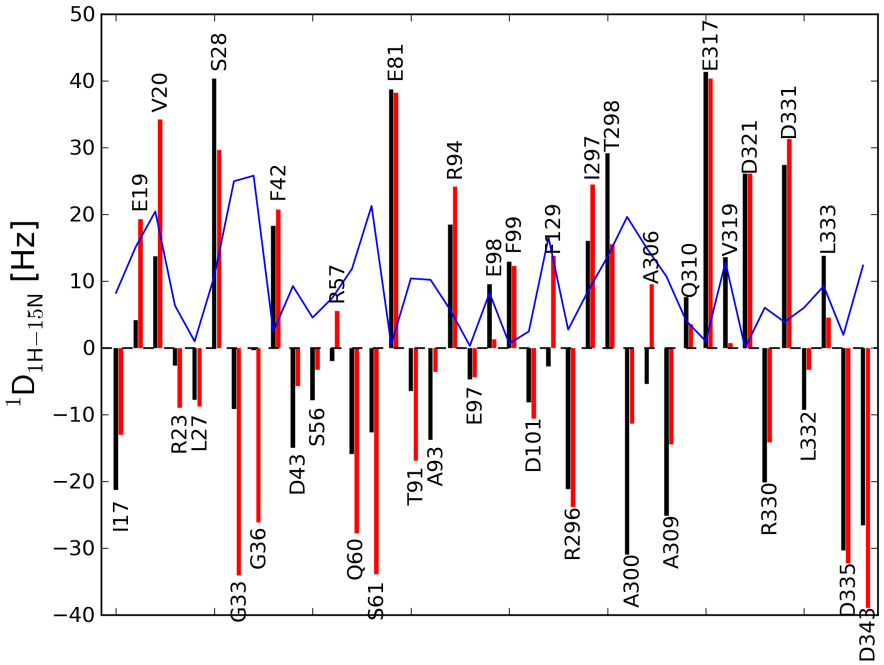


Figure S3: Direct comparison of dipolar couplings measured on free p38α (black bars) and 10b-complexed p38α (red bars). The blue line indicates the absolute difference between couplings measured for the same residue.

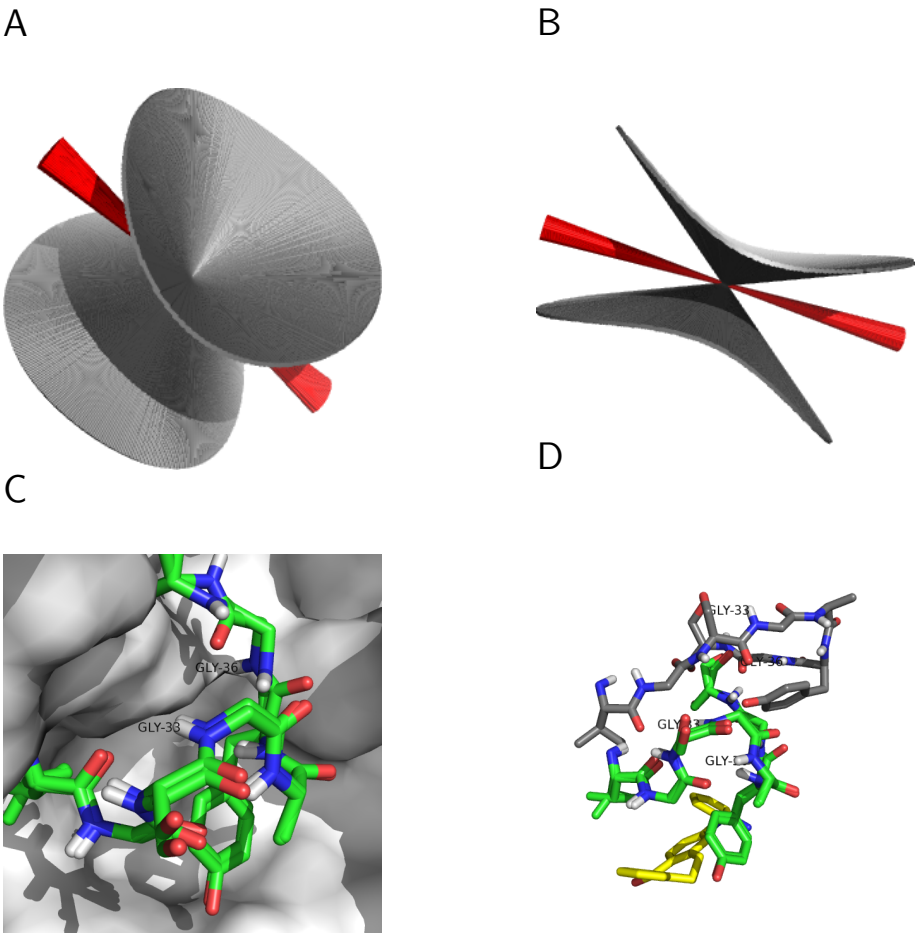


Figure S4: Optimal orientation of the N-H vectors resulting from dipolar couplings in free (grey) and inhibitor-bound (red) form. A: orientations resulting from the couplings of Gly36 which are -0.3 and -26.1 Hz for free and ligand-bound p38α, respectively. B: orientations resulting from the couplings of Gly33 which are -9.1 and -34.1 Hz, respectively. C: conformations of the glycine-rich loop in the crystal and in the hybrid structure. The amide groups of residues Gly33 and Gly36 are highlighted by labels. D: conformations of the glycine-rich loop in the crystal and hybrid structure (green carbons) and in free p38α (PDB ID: 1P38, grey carbons). The ligand 10b is shown in yellow.

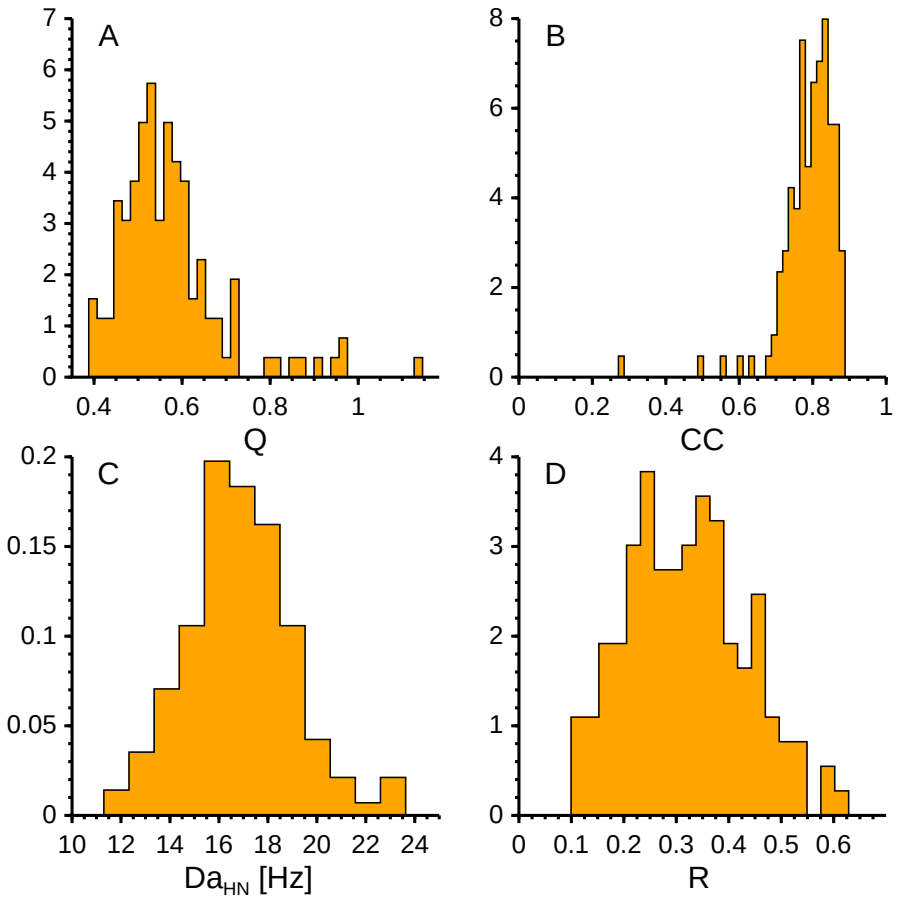


Figure S5: Survey of 138 p38α structures from PDB. (A) distribution of Q-values, the minimum value is 0.39. (B) distribution of cross-correlation between observed and calculated dipolar couplings, the maximum value is 0.89. (C,D) distribution of tensor parameters (axial and rhombic components Da_{HN} and R). Q/CC values and tensor parameters were calculated with Pales (Zweckstetter, 2008).

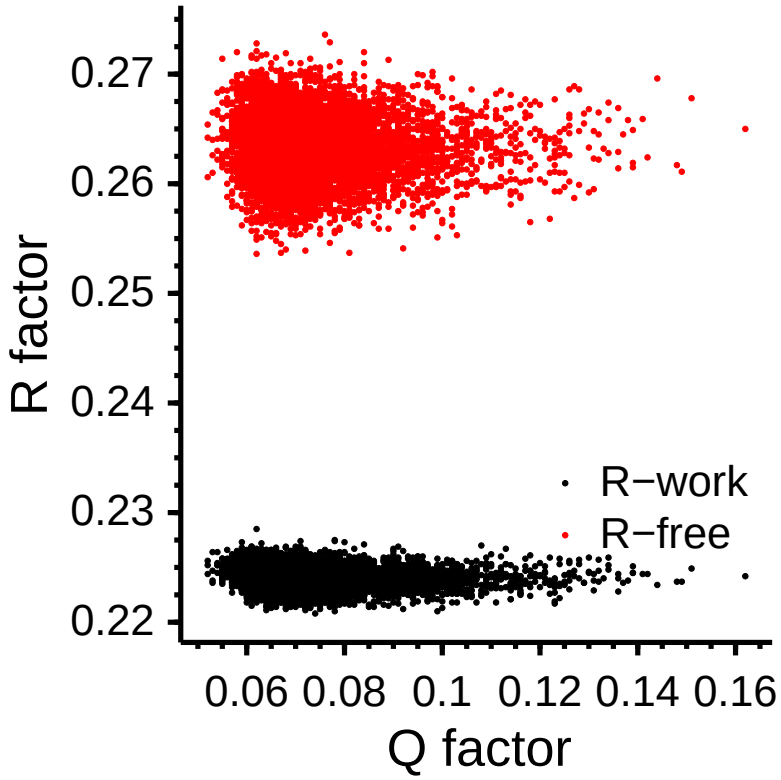


Figure S6: Correlation between NMR Q-factor and crystallographic R-factors in the final ensemble of NMR/X-ray refinement. Black dots correspond to R_{work} values, red dots indicate R_{free} values.

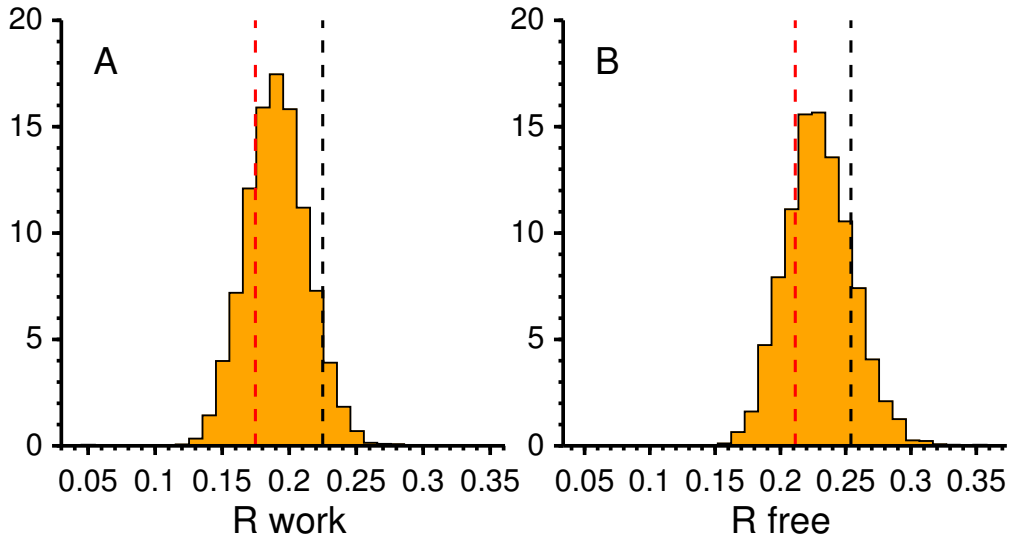


Figure S7: Distribution of R values extracted from the PDB. Shown are the distributions of R_{work} and R_{free} for structures with resolution between 1.8 and 1.9 Å; the structure by Karcher *et al.* has been determined at a resolution of 1.85 Å. The dashed red line indicates the R values of the crystal structure of p38α/10b, the dashed black lines are the R values of the hybrid structure. The structures were retrieved from the PDB by using the resolution range (1.8–1.9 Å) as sole selection criterion. 7647 structures providing full information about R_{work} and R_{free} were available and used to estimate the R value histograms.

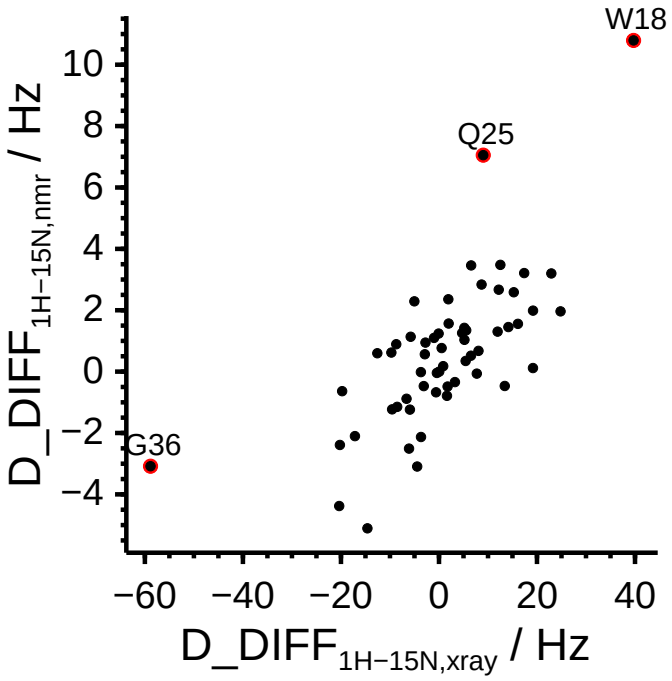


Figure S8: Discrepancy $\mathrm{D\_DIFF_{1H-15N}} = \mathrm{D\_OBS_{1H-15N}} - \mathrm{D\_CALC_{1H-15N}}$ (calculated with Pales) between observed and back-calculated dipolar couplings using the crystal structure (x-axis) and the hybrid structure of 10b/p38α (y-axis). Amino acids showing the largest discrepancies are highlighted in red.

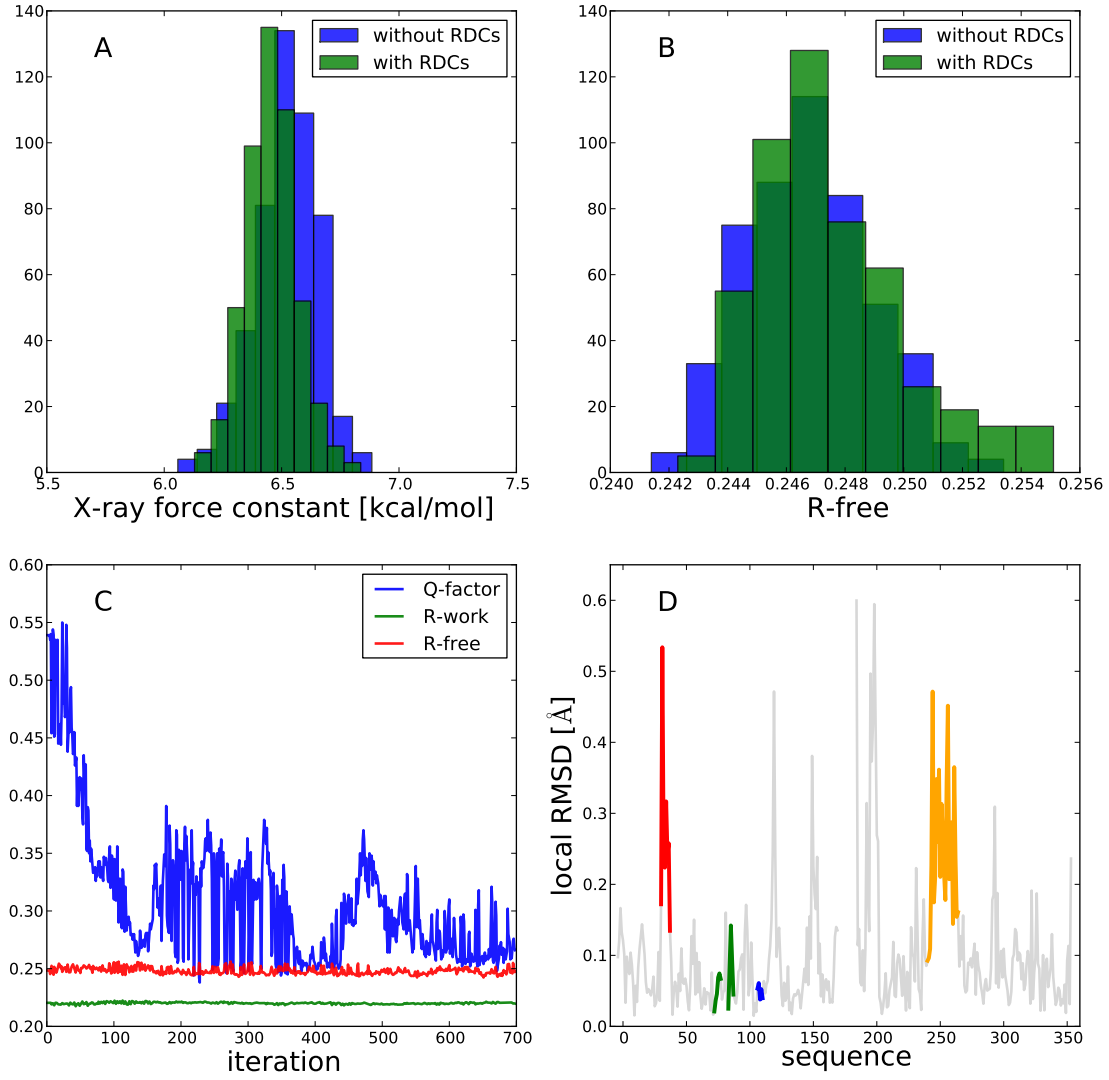


Figure S9: Refinement with adaptive X-ray force constant. Panels A, B show the distribution of the X-ray force constant and of the free R-value with and without additional RDC restraints (the RDCs were incorporated with a force constant corresponding to an average error of 2.5 Hz). Panel C shows the evolution of the Q-factor in the simulation with RDC restraints. Also shown are the R-values which stay approximately constant while the Q-factor improves. Panel D shows the local RMSD between the X-ray structure of p38α/10b and the hybrid structure (using an adaptive X-ray force constant) that has the lowest free R-value. This plot corresponds to Figure 4A in the main manuscript (red: glycine-rich loop, blue: hinge region, green: hydrophobic pocket, orange: C lobe residues).

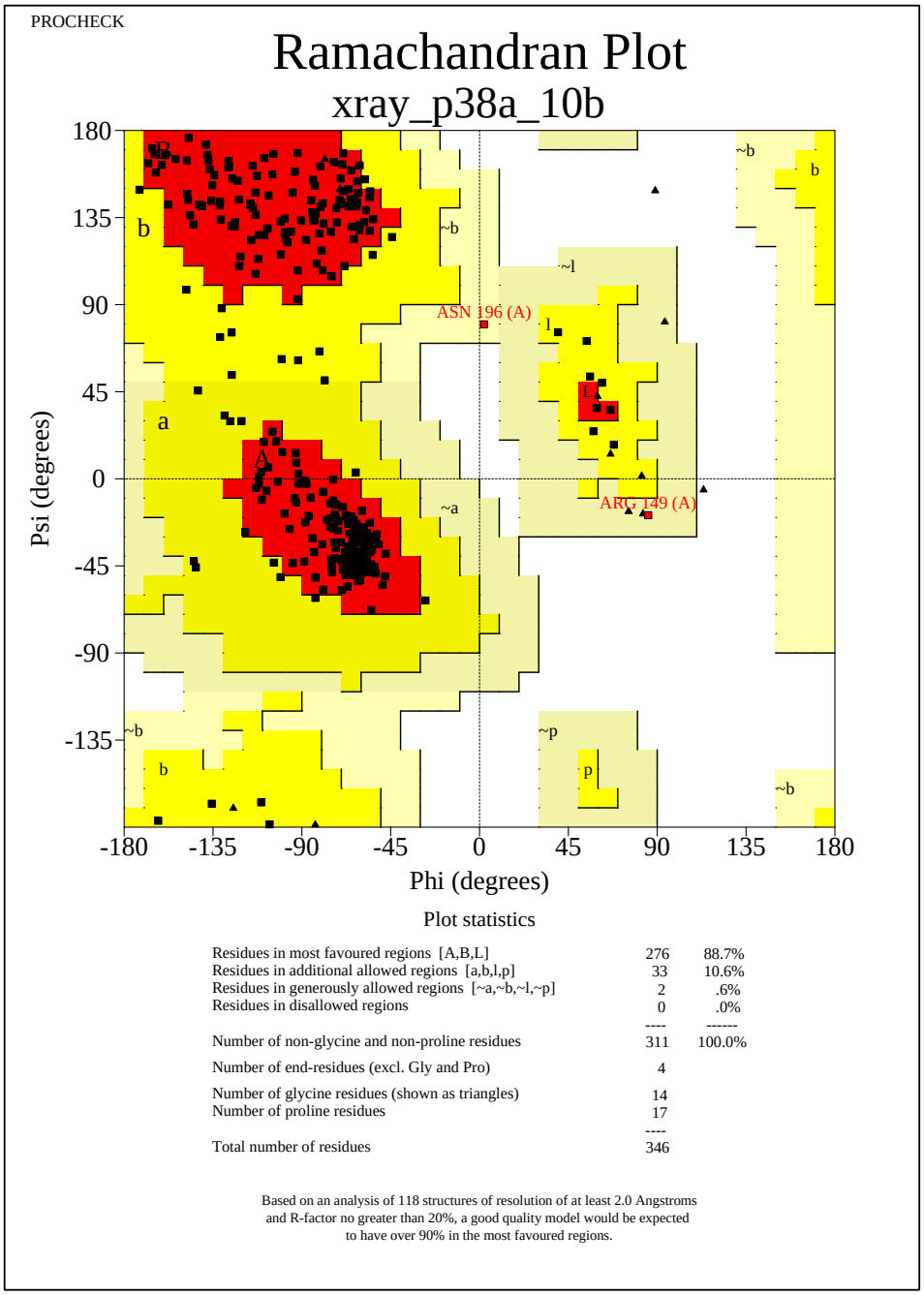


Figure S10: Ramachandran statistic of the crystal structure of 10b/p38α calculated with Procheck.

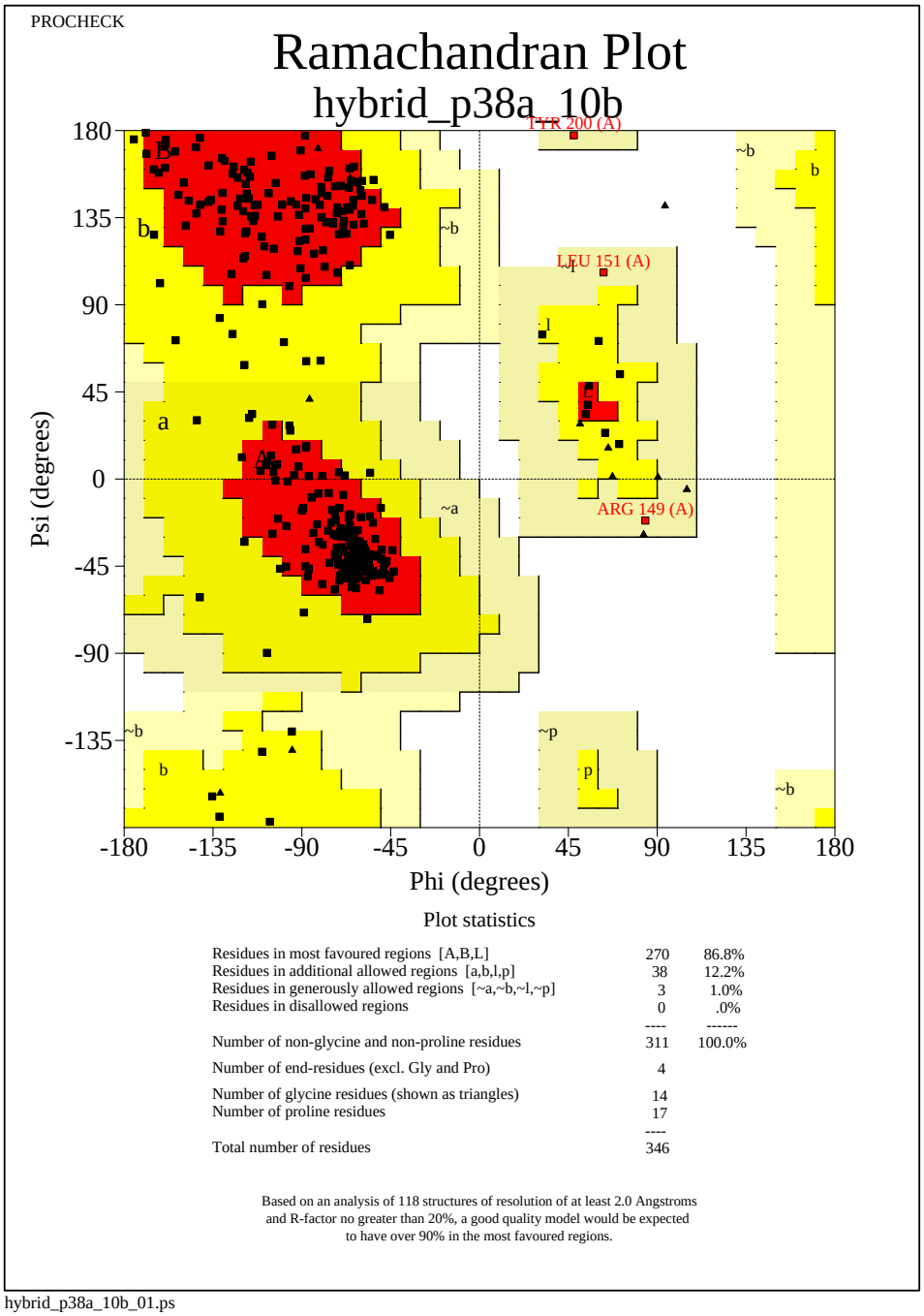
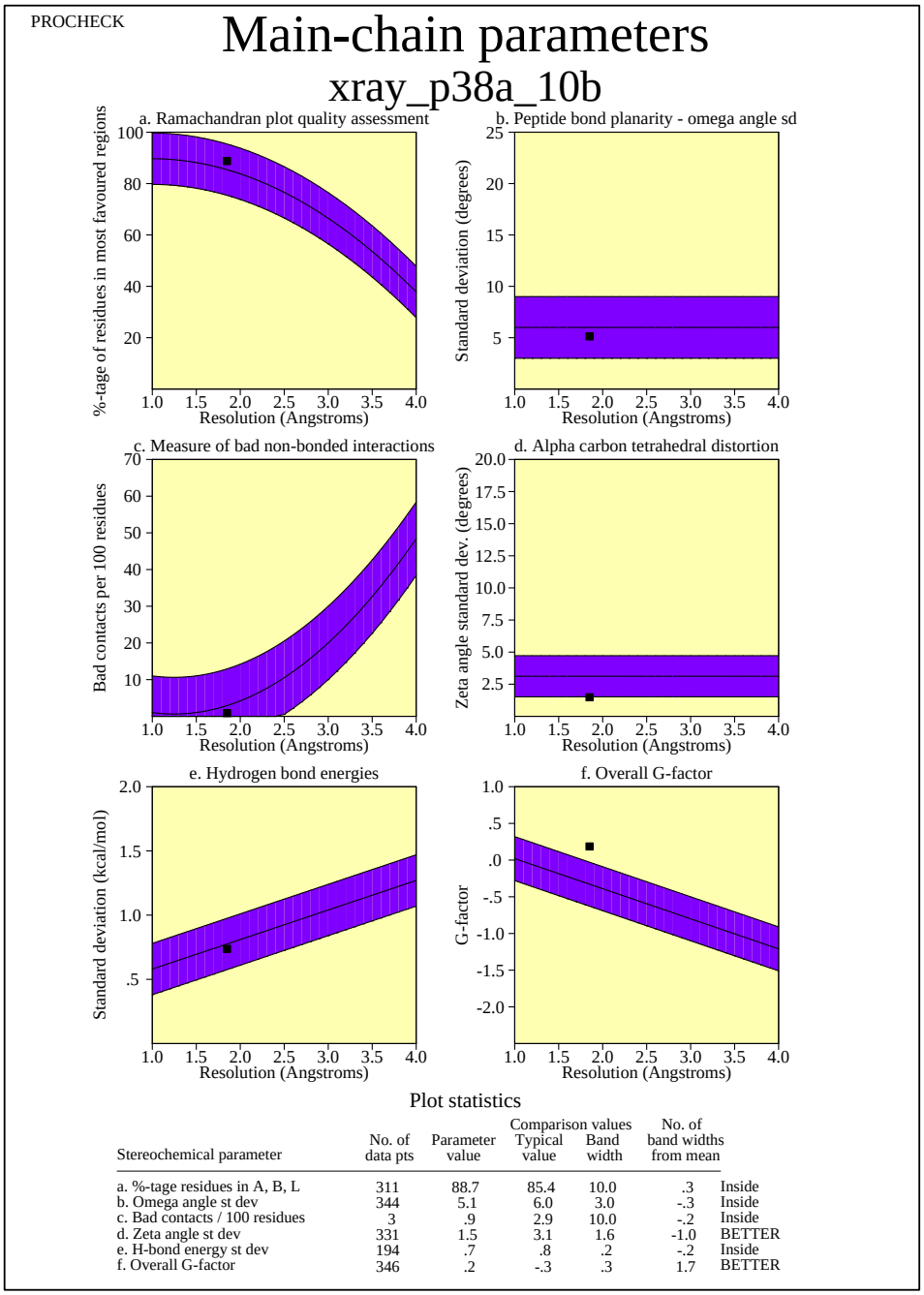
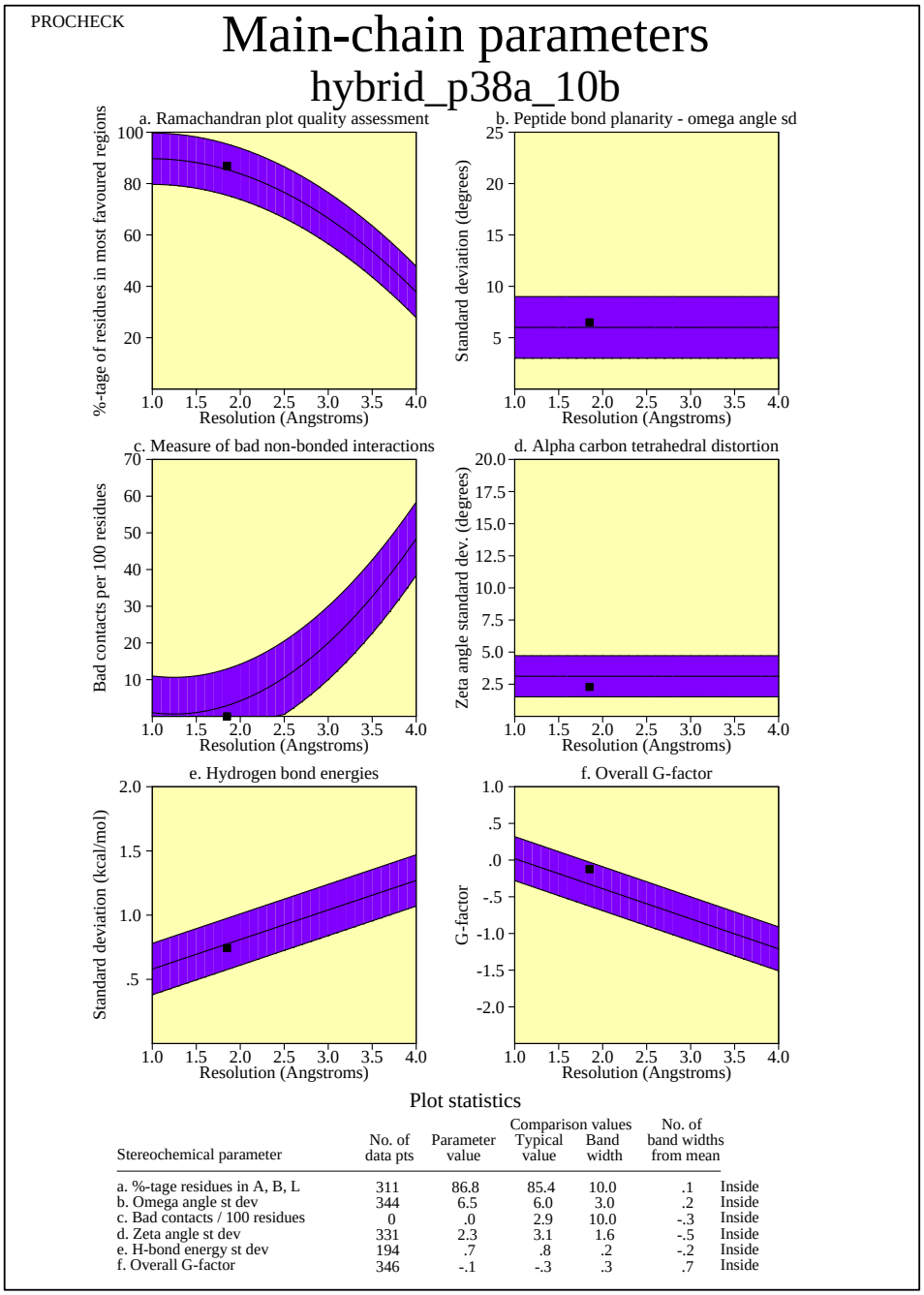


Figure S11: Ramachandran statistic of the hybrid structure of 10b/p38α calculated with Procheck.



xray_p38a_10b_04.ps

Figure S12: Main chain parameters of the crystal structure of 10b/p38α calculated with Procheck.



hybrid_p38a_10b_04.ps

Figure S13: Main chain parameters of the hybrid structure of 10b/p38α calculated with Procheck.

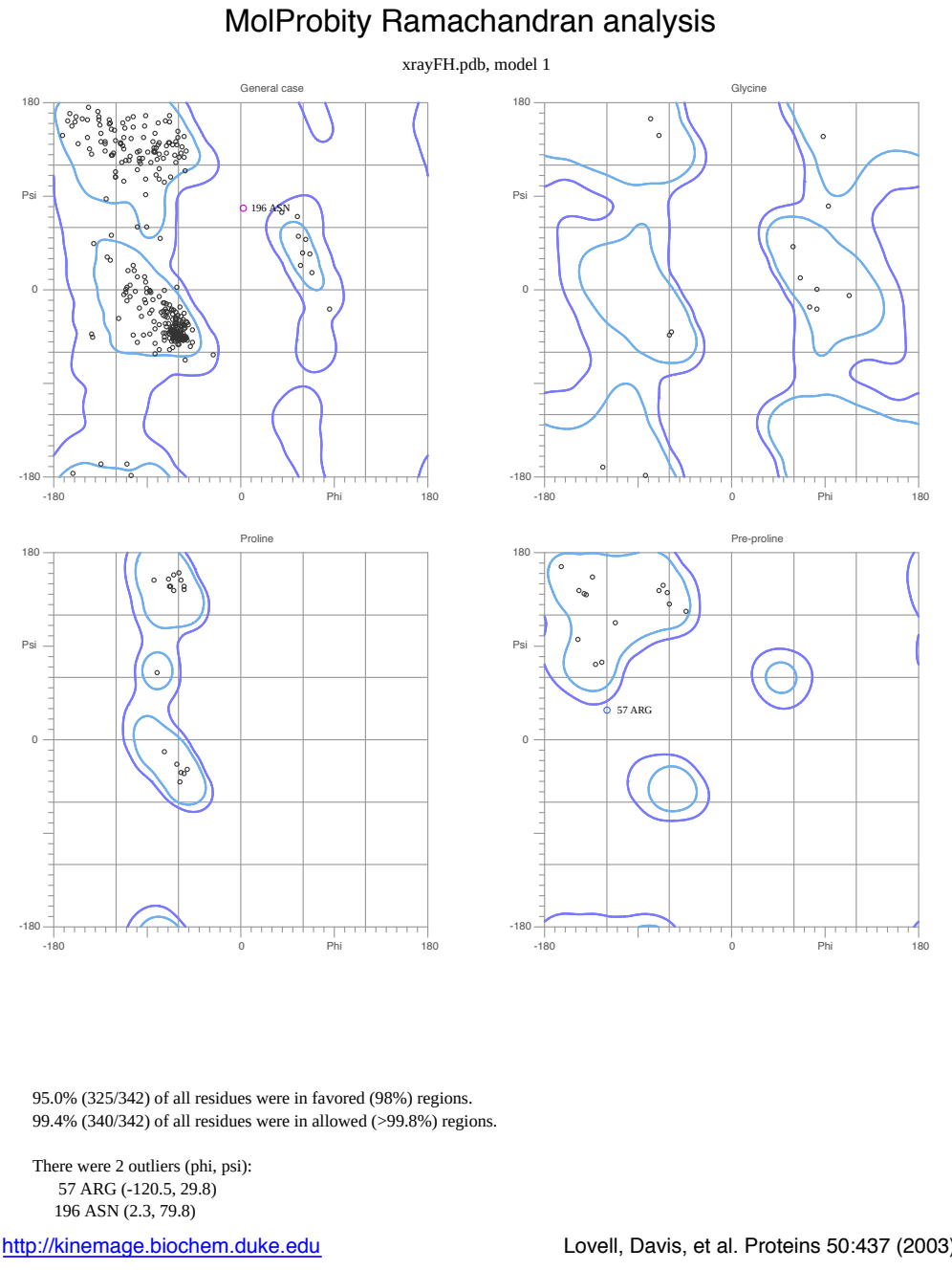
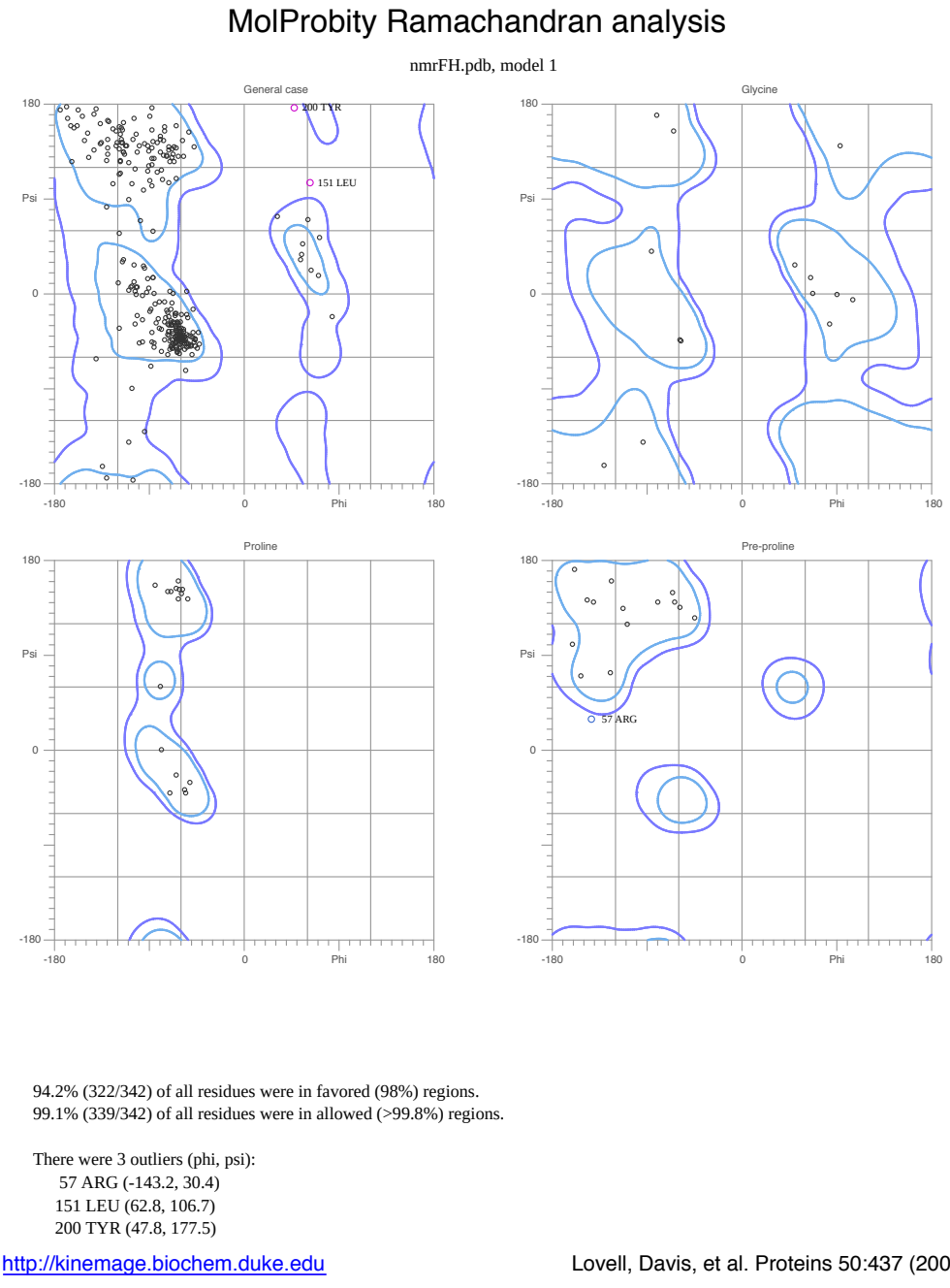


Figure S14: Molprobity Ramachandran analysis of the crystal structure of 10b/p38α.



94.2% (322/342) of all residues were in favored (98%) regions.
99.1% (339/342) of all residues were in allowed (>99.8%) regions.

There were 3 outliers (phi, psi):
57 ARG (-143.2, 30.4)
151 LEU (62.8, 106.7)
200 TYR (47.8, 177.5)

http://kinemage.biochem.duke.edu

Lovell, Davis, et al. Proteins 50:437 (2003)

Figure S15: Molprobity Ramachandran analysis of the hybrid structure of 10b/p38α.

Movie

The first movie shows the flexible fitting trajectory when refining 1P38 against the p38α/10b RDCs. During refinement, we apply two forces: (i) a force that tries to match the structure with the p38α/10b dipolar couplings, (ii) a second force based on positional restraints (Eq. (4)) that penalize deviations from 1P38 (without this force, the molecule would unfold). In the first part of the movie, the first force dominates resulting in a closure of the glycine-rich loop. In the second part of the movie, the restoring contributions of the second force partially re-open the loop. Highlighting is the same as in Fig. 2 of the manuscript (red: glycine-rich loop, blue: hinge region, green: hydrophobic back pocket).

Supplementary methods

RDC analysis

We measured $^{1}D_{1H-15N}$ residual dipolar couplings (RDCs) of the p38α/10b complex using bacteriophage Pf1 (20 mg mL^{-1}) as a weakly oriented medium. All Q and CC values reported in Table 1 and throughout the manuscript are calculated with Pales (Zweckstetter, 2008).

Comparison of RDCs from free and 10b-bound p38α

The p38α/10b RDCs were compared with those measured on free p38α (Honndorf *et al.*, 2008). Direct comparison of RDCs is only justified if the alignment is roughly the same. We checked this by fitting the crystal structure of free p38α (1P38) to the RDCs of free and 10b-bound p38α. Because not all couplings can be fitted accurately, we replaced the standard least-squares regression with a robust outlier-tolerant version involving minimization of the loss function:

$$f(\mathbf{s}) = \text{median}\{|D_i - \mathbf{a}_i^T\mathbf{s}|/\sigma_i; i = 1, \ldots, n\} \quad (1)$$

where D_i and σ_i are the ith observed RDC and its error, $\mathbf{a}_i$ a five-dimensional vector encoding the orientation of the ith N-H bond vector (Eq. (4) in Habeck *et al.* (2008)) and $\mathbf{s}$ the vector comprising the five independent tensor elements. $f(\mathbf{s})$ cannot be minimized analytically using, for example, singular-value decomposition (Losonczi *et al.*, 1999). We use numerical minimization (Powell algorithm) to find the optimal tensors. The resulting tensors are

$$\begin{aligned} S_{zz} &= 1.037 \times 10^{-3},\ S_{xx} - S_{yy} = 9.601 \times 10^{-4},\ S_{xy} = -3.127 \times 10^{-4}, \\ S_{xz} &= 9.134 \times 10^{-4},\ S_{yz} = -8.969 \times 10^{-5},\ \text{Da}_{\text{HN}} = 17.060\,\text{Hz},\ \text{R} = 0.272 \end{aligned}$$

for free p38α and

$$\begin{aligned} S_{zz} &= 8.628 \times 10^{-4},\ S_{xx} - S_{yy} = 1.004 \times 10^{-3},\ S_{xy} = -1.372 \times 10^{-4}, \\ S_{xz} &= 8.077 \times 10^{-4},\ S_{yz} = 1.379 \times 10^{-4},\ \text{Da}_{\text{HN}} = 14.756\,\text{Hz},\ \text{R} = 0.306 \end{aligned}$$

for 10b-complexed p38α, which indicates that the alignment of free and 10b-bound p38α is roughly the same.
For a given alignment tensor, the optimal orientations reproducing an observed dipolar coupling can be found by finding all orientations that minimize:

$$f(\varphi_i, \theta_i) = [D_i - A\,(3\cos^2\theta_i - 1 + 3R\sin^2\theta_i\cos(2\varphi_i))/2]^2 \qquad (2)$$

where A, R are the axial and rhombic component of the alignment tensor, φ_i and θ_i are the spherical coordinates of the ith bond vector. This results in a cone of possible bond vectors that fit the coupling optimally (Bax *et al.*, 2001).

Determination the initial alignment tensor

RDC based refinement requires an initial alignment tensor. The survey of all known p38α structures (Fig. S5) shows that no single structure is suited to define the initial tensor. From all p38α structures N-H vectors were extracted resulting in a set of candidate N-H orientations for each observed dipolar coupling. We find an alignment tensor that fits the measured couplings as best as possible by selecting, for each coupling, among the H-N orientations observed in the total set of known p38α structures. The loss function that defines the optimal tensor is the median of the discrepancies between experimental and theoretical values over all dipolar couplings:

$$g(\mathbf{s}) = \text{median}\{|D_i - D_i^{\min}(\mathbf{s})|; i = 1, \ldots, n\}. \qquad (3)$$

$D_i^{\min}(\mathbf{s})$ is the RDC resulting from a given tensor $\mathbf{s}$ and the best matching N-H orientation:

$$D_i^{\min}(\mathbf{s}) = \min\{D_{ij}(\mathbf{s}); j = 1, \ldots, 138\}$$

where $D_{ij}(\mathbf{s})$ is the ith dipolar coupling calculated with the jth p38α structure. The target function $g(\mathbf{s})$ is minimized using Powell minimization. The axial and rhombic components of the resulting tensor are 24.765 Hz and 0.545, respectively, with CC/Q = 0.99/0.098.

RDC refinement of free p38α

The structure of free p38α (PDB code: 1P38) was subjected to RDC refinement (see movie). We refine 1P38 simultaneously against the RDCs and positional restraints that restrict the conformational sampling to the vicinity (see Eq. (4) below) of the original structure. We use ISD (Rieping *et al.*, 2008) for flexible fitting and apply local sampling with Hamiltonian Monte Carlo in dihedral angles. When applying a relatively strong force ($k_{dipolar} = 10$ kcal Hz^{-2}) the fit becomes perfect within 100 refinement steps (CC/Q = 1.0/0.03).

Joint NMR/X-ray refinement

Calculation of R values and X-ray refinement of hybrid structure

Refmac5 (Murshudov *et al.*, 1997) was used to calculate R_{work}/R_{free} and to refine the hybrid structure against the structure factors obtained by Karcher *et al.*. In reciprocal space refinement, the initial structure is obtained by replacing the coordinates of p38α in the original structure of p38α/10b (Karcher *et al.*) with those obtained with ISD during NMR/X-ray refinement. Coordinates of the ligand and water remain unchanged. Refinement was done with default parameters.

NMR/X-ray refinement with ISD

Because joint structure factor and RDC refinement is not possible with Refmac5, we use ISD (Rieping *et al.*, 2005, 2008) to sample structures that fit both NMR and X-ray data. A crystallographic R-factor is not implemented in ISD and its evaluation would be too inefficient for exhaustive conformational sampling. To mimic an R-factor term, we apply positional restraints for all atoms that have coordinates in the crystal structure. The strength of each restraint was chosen in correspondence with its B-factor: atoms that are more mobile are also allowed to vary to a greater extent in the refinement. The X-ray restraint energy

$$E_{\mathrm{xray}}(\theta) = \sum_i \frac{8\pi^2}{B_i} \|\mathbf{x}_i - \mathbf{x}_i(\theta)\|^2 \tag{4}$$

was implemented in ISD. Here, θ are the dihedral angles of p38α that serve as only conformational degrees of freedom during conformational sampling (Habeck *et al.*, 2005b), B_i is the temperature factor of the ith heavy atom, $\mathbf{x}_i$ its position in the crystal or Refmac5-refined structure. To calculate a structure that fits both crystallographic and dipolar coupling data, we added the standard least-squares residual for dipolar coupling data (Habeck *et al.*, 2008) to the X-ray energy. Conformational sampling was carried out using replica exchange Monte Carlo in which the strength of the forcefield and of the positional restraints served as fictitious temperatures (Habeck *et al.*, 2005b).

Calculation of the initial structure for first round of refinement

The Modeller program (Sali and Blundell, 1993) was used to model a complete structure of p38α/10b. The crystal structure of p38α/10b (Karcher *et al.*) served as template, residues that could not be built originally (especially residues 171–180) were added and optimized by Modeller. NMR/X-ray refinement was applied to the homology model. A fictitious temperature λ ranging from 1 to 0.1 is applied to the positional restraints only. First, the PROLSQ force field as implemented in ISD serves as conformational prior (Habeck *et al.*, 2005a). The alignment tensor is held fixed, but the force constant is estimated (Habeck *et al.*, 2008). From the replica with $\lambda = 0.1$, we selected a promising candidate for further refinement that showed a good Q-value (Q = 0.098) as well as an acceptable alpha-carbon RMSD of 0.8 Å to the crystal structure by Karcher *et al.*. This structure was further refined using the Rosetta non-bonded force field (Kuhlman *et al.*, 2003) with fixed tensor parameters. From the replica with $\lambda = 1.0$, a structure with good CC/Q values and low RMSD to the crystal structure was selected. This structure fits the RDCs well (CC/Q = 0.96/0.19) but still has poor R values R_{work}/R_{free} = 0.346/0.359. Crystallographic refinement with Refmac5 improves the R values to R_{work}/R_{free} = 0.199/0.248 at the expense of deteriorating the fit to the NMR data (CC/Q = 0.82/0.56). In a subsequent short replica simulation (560 transitions) with fixed RDC force constant ($k_{dipolar}$ = 1 kcal Hz^{-2}) using the PROLSQ forcefield, we obtained a structure with CC/Q = 0.96/0.16 and

R_{work}/R_{free} = 0.28/0.30. This structure served as *initial structure* for the first round of joint NMR/X-ray refinement and was used as reference in the definition of the X-ray residual (Eq. (4)).

Two rounds of NMR/X-ray refinement

Throughout NMR/X-ray refinement we use the PROLSQ forcefield as conformational prior, the RDC force constant is fixed ($k_{dipolar}$ = 1 kcal Hz^{-2}) but the alignment tensor is allowed to vary. Initially, the refinement starts with a structure of moderate quality (*initial structure* with CC/Q = 0.96/0.16, R_{work}/R_{free} = 0.28/0.30). The X-ray residual is then optimized at the expense of diminishing the fit to the residual dipolar couplings, which is reflected in an increase in the NMR restraint energy. After 100 replica iterations, the ISD samples conformations that improve in both restraint energies conjointly. During this round of refinement the structure improves to CC/Q = 0.99/0.11 and R_{work}/R_{free} = 0.23/0.27. After crystallographic refinement of the first hybrid structure using Refmac5, we update the reference in the X-ray residual (Eq. (4)) which is then used in a second round of refinement. This round of refinement improves the hybrid structure further to CC/Q = 0.995/0.062 and R_{work}/R_{free} = 0.225/0.254. The estimated alignment tensor is Da_{HN} = 26.278 Hz and R = 0.446. Figure S6 shows the quality factors of the sampled p38α conformations.

Joint refinement with adaptive force constant

The hybrid structure obtained by joint X-ray/NMR refinement is based on a fixed RDC force constant that corresponds to an average RDC error of 1 Hz. Also the force constant of the X-ray residual was fixed to a constant (1 kcal/mol). This produced a hybrid structure with reasonably high R-values but unrealistically low Q-factors. To obtain a hybrid structure that shows a more realistic fit with the RDCs, we ran a simulation in which the RDC force constant $k_{dipolar}$ was fixed to a value that corresponds to the average error of the 58 RDCs (2.5 Hz). Furthermore, to balance the strength of X-ray residual relative to the RDC restraints we sampled the weight of the X-ray term using Bayesian inference (Habeck *et al.*, 2006). We

ran two simulations: one without RDCs, one with RDCs. Figure S9A shows that the estimated force constant of the X-ray residual is not affected by the inclusion of the RDCs and that it reaches a value of 6.52 kcal/mol (without RDCs) and 6.45 kcal/mol (with RDCs). This again shows that the RDC data do not contradict the X-ray data. Also the distribution of the free R-value does not change significantly upon addition of the RDCs (Fig. S9B). However, as shown in Figure S9C the Q-factor improves from 0.55 to 0.25, if we include the RDCs in the simulation using a force constant that corresponds to an average error of 2.5 Hz. The best Q-factor is 0.238 with a corresponding $R_{work}/R_{free} = 0.22/0.25$. The best R-values are $R_{work}/R_{free} = 0.22/0.24$ with a corresponding Q-factor of 0.254. The structure with the best R-values is highly similar to the hybrid structure obtained with fixed force constants (see previous section). Figure S9D shows the local RMSD with the crystal structure of p38α/10b. Comparison with Figure 4A of the main manuscript illustrates that both hybrid structures convey the same information.

Validation

We checked the quality of the final structure ensemble (PDB code: 2LGC) using Procheck and Molprobity. In the crystal structure of p38α/10b, 88.7% residues are in favoured, 10.6% in allowed, 0.6% in generously allowed, and 0.0% in disallowed regions (Figure S10). In the hybrid structure, 86.8% residues are in favoured, 12.2% in allowed, 1.0% in generously allowed, and 0.0% in disallowed regions (Figure S11). Procheck finds two suspicious residues for the crystal structure (Arg149, Asn196), and three for the hybrid structure (Arg149, Leu151, Tyr200). Molprobity gives similar results but labels different residues as outliers (Figures S14 and S15). The crystal structure has 95.0% in favoured (98%) regions, and 99.4% in allowed ($>$ 99.8%) regions. The hybrid structure has 94.2% in favoured (98%) regions, and 99.1% in allowed ($>$ 99.8%) regions. Molprobity considers different residues as outliers. For the crystal structure it marks Arg57 and Asn196 as outliers; for the hybrid structure it finds Arg57, Leu151, Tyr200. Asn196 (outlier in the crystal structure) as well as Leu151 and Tyr200 (outliers in the hybrid structure) are all in the vicinity of the activation loop 180-188, which is an unstructured region and

almost completely disordered or missing in the crystal structure. According to Procheck, the hybrid structure is more regular with an Overall G-factor of -0.1 than the crystal structure (G-factor = 0.2); the typical G-factor at resolution 1.85 Å is -0.3 (Figures S12 and S13).

In the hybrid refinement, the structures of the individual lobes did not change much. This is expected because of the rigid arrangement of secondary structure elements. This implies that the RDC-derived Q factors obtained for N and C lobe separately should be smaller than the Q factor of the entire kinase compared to the crystal or the hybrid structure, respectively. To test whether the RDC data reflect these observations, we carried out several analyses for reduced sets of RDCs using the crystal structure of p38α/10b and the crystal structure of free p38α (1P38). For the crystal structure of p38α/10b, the Q factor achieved with RDCs of p38/10b in secondary structure only is 0.39 overall, 0.47 for the N lobe and 0.26 for the C lobe. For the free structure of p38α (1P38) we obtain Q factors of 0.49 (overall), 0.58 (N lobe), and 0.32 (C lobe). If we remove the outliers in the second beta strand (residues 18, 19, 20), the fits obtained with the crystal structure of the complex improve to 0.23 (overall), 0.20 (N lobe), and 0.26 (C lobe). For the free structure of p38α (1P38), the corresponding values are: 0.28 (overall), 0.25 (N lobe), and 0.30 (C lobe). These numbers show that the RDCs, which are part of the core, can be described better with the structure of the complex than with the free structure. The numbers also show that after removal of the outliers the fits obtained with the complex are reasonably good. However, the numbers also show that the core of the individual lobes does not change dramatically, since the cores of the N and C lobes of the free structure fit almost as well to the RDCs as the p38α/10b complex structure. The fit achieved for the residues in secondary structure elements is acceptable for such a large protein and supports the idea that the core of the protein is well defined, does not change dramatically for the cores of the two lobes and is consistent with the RDC data.

References

Bax, A., Kontaxis, G., and Tjandra, N. (2001). Dipolar Couplings in Macromolecular Structure Determination. *Methods in Enzymology*, **339**, 127–174.

Habeck, M., Nilges, M., and Rieping, W. (2005a). Bayesian inference applied to macromolecular structure determination. *Phys. Rev. E*, **72**, 031912.

Habeck, M., Nilges, M., and Rieping, W. (2005b). Replica-Exchange Monte Carlo scheme for Bayesian data analysis. *Phys. Rev. Lett.*, **94**, 0181051–0181054.

Habeck, M., Rieping, W., and Nilges, M. (2006). Weighting of experimental evidence in macromolecular structure determination. *Proc. Natl. Acad. Sci. USA*, **103**, 1756–1761.

Habeck, M., Nilges, M., and Rieping, W. (2008). A unifying probabilistic framework for analyzing residual dipolar couplings. *J. Biomol. NMR*, **40**, 135–144.

Honndorf, V. S., Coudevylle, N., Laufer, S., Becker, S., and Griesinger, C. (2008). Dynamics in the p38alpha MAP kinase-SB203580 complex observed by liquid-state NMR spectroscopy. *Angew. Chem. Int. Ed. Engl.*, **47**, 3548–3551.

Kuhlman, B., Dantas, G., Ireton, G. C., Varani, G., Stoddard, B. L., and Baker, D. (2003). Design of a novel globular protein fold with atomic-level accuracy. *Science*, **302**, 1364–1368.

Losonczi, J. A., Andrec, M., Fischer, M. W. F., and Prestegard, J. H. (1999). Order Matrix Analysis of Residual Dipolar Couplings Using Singular Value Decomposition. *J. Magn. Reson.*, **138**, 334–342.

Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.*, **53**, 240–255.

Rieping, W., Habeck, M., and Nilges, M. (2005). Inferential Structure Determination. *Science*, **309**, 303–306.

Rieping, W., Nilges, M., and Habeck, M. (2008). ISD: a software package for Bayesian NMR structure calculation. *Bioinformatics*, **24**, 1104–1105.

Sali, A. and Blundell, T. (1993). Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.*, **234**, 779–815.

Zweckstetter, M. (2008). NMR: prediction of molecular alignment from structure using the PALES software. *Nat Protoc*, **3**, 679–690.