Supplementary Material:
Conformational selection by the aIF2 GTPase: a molecular dynamics study of functional pathways

Priyadarshi Satpati and Thomas Simonson
Laboratoire de Biochimie (CNRS UMR7654), Department of Biology,
Ecole Polytechnique, 91128 Palaiseau, France.

In addition to the material provided here, three MD structures are also available as separate files, corresponding to the aIF2:GDP:Pi complexes in the three different states.

pK_a calculations for the ligands and His97: methods and results

Free energy calculations were used to determine the protonation state of the GDP and Pi ligands and of the nearby His97 in the MIXED state. We use a "Poisson-Boltzmann Linear Response Approximation", or PB/LRA method [1–4]. Protonation of a specific group (phosphate group, His sidechain) is modelled by changing selected atomic charges. The corresponding free energy change is computed both in the protein complex and for the same group alone in solution, using conformations generated by MD simulation. The free energy changes are approximated by the continuum electrostatic free energy, where the protein and ligand atoms are explicitly included but the solvent is replaced by a dielectric continuum. The free energy change in either medium (protein, solution) can be written:

$$\Delta G = \frac{1}{2}\sum_i \delta q_i (V_i^A + V_i^B). \quad (1)$$

The sum is over all the atoms of the titrable group; δq_i is the change in the atomic charge due to the protonation; V_i^A (respectively, V_i^B) is the electrostatic potential on atom i when the proton charge is absent (respectively, present). These potentials are averaged over an MD simulation performed with the proton charge absent (A) or present (B). The protein was thus simulated in explicit water in the two states of interest (proton absent/present). The electrostatic potentials were then computed for each MD conformation, by numerically solving the Poisson equation of continuum electrostatics, where the protein and ligand were treated as a single dielectric medium with a dielectric constant of 1 or 2; solvent was treated as another medium with a dielectric constant of 80. A low dielectric value is appropriate for the protein because the conformational changes induced by the protonation reaction are explicitly modelled, through the use of

the two endpoint MD simulations [1–4]. The boundary between the two dielectric media was defined as the protein/ligand molecular surface, computed with a 2 Å radius probe sphere. For the potential calculation, the system was discretized using a cubic grid with a 130 Å edge and a spacing of 0.4 Å. The Poisson equation was solved numerically, with Coulombic boundary conditions, using the Charmm program (PBEQ module) [5]. A physiological ionic strength of 0.1 M was used. The same procedure was followed for the same titrable group in water, using the conformations sampled in the protein simulations (but discarding the protein atoms). The pK_a shift due to the protein environment has the form:

$$pK_{a,prot} - pK_{a,solv} = \frac{1}{2.303 k_B T}(\Delta G_{prot} - \Delta G_{solv}), \tag{2}$$

where k_B is Boltzmann's constant and T is the temperature. Protonation state calculations for GDP and Pi were performed for 189 MD conformations. For His97, calculations were performed for 400 conformations.

Results are given in Table 1. Pi protonation, to form $H_2PO_4^-$, is much harder in the protein than in solution, regardless of the assumptions made for GDP and His97. Thus, we will model Pi as singly protonated, with a net charge of -2. For GDP, protonation to form $GDPH^{2-}$ is very unfavorable in the protein, compared to solution. Thus, we model GDP as fully deprotonated, with a net charge of -3. As a consistency check for these results, we may consider two ways to convert the GDP^{3-}:HPO_4^{2-} complex into the $GDPH^{2-}$:$H_2PO_4^-$ complex: by protonating either GDP or Pi first. The results (with $\epsilon_p = 2$) differ by 2.9 kcal/mol, which provides a rough error estimate, and is comparable to the estimated statistical errors for the individual pK_a shifts (Table 1). The estimated error is significant, but much smaller than the computed pK_a shifts. Finally, with the ligands GDP^{3-} and HPO_4^{2-}, His97 prefers to be doubly protonated and positively charged. In this case, the preference is not far from the statistical error, 2.5 kcal/mol when using a protein dielectric of two. Fortunately, the free energy for Pi dissociation is not sensitive to the His97 protonation state (see Table 2, below). This observation indicates that the His97 pK_a does not change when Pi dissociates, so that our MD model with a doubly-protonated His97 is also valid for the MIXED:GDP complex.

Table 1: Protonation state calculations for GDP, Pi, and His97 in the MIXED state

$HPO_4^{2-} \rightarrow H_2PO_4^-$					
Ligand	ϵ_p	$\Delta G_{\rm soln}$	$\Delta G_{\rm prot}$	$\Delta\Delta G$	pK_a shift
GDP^{3-}	1	-64.1	-52.8	11.1 (1.9)	-8.1
GDP^{3-}	2	-34.5	-28.1	6.5 (0.8)	-4.7
$GDPH^{2-}$	1	-65.5	-53.3	12.2 (2.4)	-8.9
$GDPH^{2-}$	2	-35.3	-27.8	7.5 (0.6)	-5.5
$GDP^{3-} \rightarrow GDPH^{2-}$					
Ligand	ϵ_p	$\Delta G_{\rm soln}$	$\Delta G_{\rm prot}$	$\Delta\Delta G$	pK_a shift
HPO_4^{2-}	1	-1.6	30.1	31.8 (5.0)	-23.1
HPO_4^{2-}	2	-4.2	14.2	18.5 (3.1)	-13.4
$H_2PO_4^-$	1	-7.1	23.0	30.1 (1.6)	-21.9
$H_2PO_4^-$	2	-6.9	10.6	16.6 (0.8)	-12.8
His97 $\rightarrow$ $His97^+$					
Ligands	ϵ_p	$\Delta G_{\rm soln}$	$\Delta G_{\rm prot}$	$\Delta\Delta G$	pK_a shift
GDP^{3-}	1	56.3	51.8	-4.5 (0.1)	+3.3
and HPO_4^{2-}	2	28.8	26.3	-2.5 (0.1)	+1.8

Free energies (kcal/mol) to protonate a specific group: Pi (upper part), GDP (middle), or His97 (bottom), either in the protein ($\Delta G_{\rm prot}$) or in solution ($\Delta G_{\rm soln}$). Protonation of either ligand (GDP or Pi) is done with the other ligand in a particular protonation state, indicated in the first column. A protein dielectric constant ϵ_p of 1 or 2 is used, as indicated. $\Delta\Delta G = \Delta G_{\rm prot} - \Delta G_{\rm soln}$. The pK_a shift for the specific group in the protein is given relative to the same group in solution. Uncertainty, in parentheses, is estimated as the difference between the results from the first and second half of each ensemble of structures.

MD free energy runs and their uncertainty

The individual MD free energy runs simulating Pi removal are listed in Table 2, below. The total simulation length (not including solution runs) was about 330 nanoseconds. The statistical uncertainties of the individual runs (estimated by comparing the two halves of each window), range from 0.4 to 2.5 kcal/mol (Table 2). The standard deviations between individual runs are comparable: 2.6 kcal/mol for the four ON state runs, 1.4 kcal/mol for the four OFF state runs, and 1.7 kcal/mol for the MIXED runs 9 and 11 (the runs used for averaging; see below). Thus, statistical errors are small. In addition, for some of the runs, we tested the effect of using alternative models, as described next.

Runs with alternative modelling procedures For the ON and OFF complexes with GDP + Pi, we created our main models by cutting the β–γ phosphate bond of GTP, as described in the main text. An alternate procedure was used for some runs, to test the sensitivity of the free energy to the initial Pi placement. The alternate procedure started from an MD simulation of the ON:GDP or OFF:GDP complex. From each simulation, we computed the electrostatic potential at the position of each water oxygen within 10 Å of the Mg^{2+} ion, averaging over the last 500 ps of dynamics. We then positioned the Pi such that one of its negatively-charged oxygens is at the position with the most positive potential (replacing the corresponding water molecule). This procedure was used for runs 3, 4, 7, 8 in Table 2. The average results with the alternative models are quite similar to those with the main models, within about 1 kcal/mol.

For the MIXED state, most of the runs modelled His97 as doubly-protonated. This is the most stable state for the GDP + Pi complex, according to the PB/LRA calculations (Table 1). However, the computed preference is not much greater than the PB/LRA uncertainty, and so two runs were performed with a neutral His97 (runs 13, 14 in Table 2), to test the sensitivity of the Pi dissociation free energy to this assumption. As seen in Table 2, the results are very similar: *e.g.*, 402.7 for the forward run 13 versus 405.7 and 402.3 for the forward runs 9 and 11. Another possibility is for His97 to be doubly-protonated with bound GDP + Pi, but neutral after Pi dissociates. In that case, there would be another free energy contribution, related to proton dissociation in the MIXED:GDP complex. This contribution depends on the His97 pK_a in the MIXED:GDP complex, which was not calculated precisely. However, since the Pi dissociation free energy is insensitive to the His97 protonation state, the Pi is not responsible for shifting its pK_a, and the protonation state should be the same after Pi dissociation.

Comparing forward and backward runs For the MIXED state, we did runs in both directions, forward and backward. For the 13/14 pair of runs, which have a neutral His97, the forward/backward agreement is excellent. For the other runs, there is a large forward/backward discrepancy, or hysteresis (runs 9–12). Thus, the results of runs 11 and 12 differ by 8.7 kcal/mol. The free energy derivative is plotted in Figure 1 for runs 9–12, illustrating the level of agreement between runs. Agreement is quite good for the derivatives, but the total free energies are so large that small relative errors remain significant. The forward/backward hysteresis in runs 9–12 can be traced to the Lys48 sidechain, which interacts with Pi in the starting complex, along with His37 and His97. When Pi is deleted, the repulsion between these positive sidechains pushes Lys48 out of the binding pocket, as shown in Fig. 3 of the main text. This motion also occurs in the "neutral His97" runs, 13-14. When Pi is reintroduced in the backwards runs 10 and 12, Lys48 does not move back into its initial position, but remains trapped in its outwards orientation. In contrast, when His97 is neutral, Lys48 is able to return to its initial position during the backwards run. Thus, the hysteresis seen in runs 9–12 does not arise from a general "inertia" of the system, but from an incorrect positioning of Lys48 at the end of runs 10 and 12. As a result, for the MIXED state Pi dissociation free energy, we use the average over the forward runs 9 and 11. The excellent reversibility of runs 13/14 actually supports the use of only forward runs for the ON and OFF states. Notice that the run lengths for the electrostatic stage for these two states were chosen to be over three times as long as the reversible MIXED state runs: 22 vs 6.6 ns.

Other sources of error Statistical error, structural hysteresis, and assumptions about His97 protonation and Pi placement lead to small errors and uncertainty, as shown above. Two other potential sources of error are the method to integrate the free energy derivatives and the use of periodic boundary conditions with PME electrostatics. For the electrostatics stages and most of the van der Waals stages, trapezoidal integration should give good accuracy, given the smooth variation of the derivatives (Fig. 3). For the van der Waals stage close to $\lambda_{\mathrm{vdW}} = 0$, we fitted the derivative by a function ($A\lambda_{\mathrm{vdW}}^{-B}$) that ressembles the infinite branch of the free energy derivative of a vanishing (uncharged) atom in water, $A\lambda_{\mathrm{vdW}}^{-3/4}$ [6]. The contribution of this branch to the overall free energy change is only 0.56±0.07 kcal/mol, so that any uncertainty associated with this fitting procedure is very small. A good alternative would have been to use a so-called "soft sphere" van der Waals potential for the vanishing Pi group. While this procedure is known to give improved efficiency and precision [7, 8], the method used here is sufficient

for our purposes.

While periodic simulation models are one of the best methods for MDFE, they do produce artefacts and systematic errors [9]. One error arises from direct interactions between the solutes in different boxes. Thus, our divalent Pi interacts with an array of its own images; however the electrostatic free energy due to these interactions is only about 0.1 kcal/mol for the large box sizes used here; see [10] for a precise estimation. A more complex effect is the change in solvent polarization near the box edges, due to the artifical presence of identical, neighboring boxes (instead of bulk solvent). While this effect is hard to estimate, we showed earlier that when a large, distant, solvent region is replaced by protein, the effect on GTP and GDP binding is very small [10]. Also, a dielectric continuum free energy method led to very similar GTP/GDP binding free energy differences [10], suggesting that our results are robust with respect to the exact electrostatic treatment.

Finally, systematic errors due to the force field are another difficulty, which was discussed at length in an earlier study of the present system [10]. This is a complex problem, and the exact error is hard to estimate precisely. Nevertheless, our earlier work [10], as well as ongoing work (manuscript in preparation), suggest that the present results are reasonably robust with respect to the choice of force field, with the Charmm22, Amber, and polarizable AMOEBA force fields giving fairly similar results for a simplified GTP $\rightarrow$ GDP transformation.

Table 2: Individual free energy runs for HPO_4^{2-} deletion

run number	system	run direction	run length (ns)	ΔG^{elec}	ΔG^{vdw}	ΔG
1	ON	forward	22.0 + 7.8	421.5 (2.1)	-4.0 (0.6)	417.4 (2.2)
2	ON	forward	22.0 + 7.8	425.0 (1.6)	-3.1 (0.6)	421.9 (1.7)
3	ON[a]	forward	22.0 + 7.8	422.3 (0.4)	-2.2 (1.0)	420.1 (1.1)
4	ON[a]	forward	13.2 + 7.8	422.1 (0.4)	-5.9 (0.5)	416.2 (0.6)
5	OFF	forward	22.0 + 7.8	418.4 (0.2)	-2.6 (0.6)	415.8 (0.6)
6	OFF	forward	22.0 + 7.8	417.0 (0.4)	-1.7 (0.2)	415.4 (0.4)
7	OFF[a]	forward	22.0 + 7.8	416.3 (0.2)	2.1 (0.6)	418.4 (0.6)
8	OFF[a]	forward	13.2 + 7.8	419.3 (1.0)	-3.9 (0.8)	415.4 (1.3)
	solution[b]	forward	6.6 + 11.4	443.0 (0.6)	-3.5 (0.4)	439.5 (0.7)
9	MIXED	forward	6.6 + 15.8	404.3 (2.5)	1.4 (0.5)	405.7 (2.5)
10	MIXED	backward	6.6 + 7.8	390.9 (0.5)	2.2 (1.8)	393.2 (1.9)
11	MIXED	forward	13.2 + 7.8	402.3 (0.7)	-0.0 (1.5)	402.3 (1.7)
12	MIXED	backward	13.2 + 9.8	394.7 (0.3)	-1.2 (1.1)	393.6 (1.2)
13	MIXED[c]	forward	6.6 + 7.8	404.8 (0.8)	-2.1 (1.4)	402.7 (1.6)
14	MIXED[c]	backward	6.6 + 7.8	402.2 (0.5)	-1.6 (0.3)	400.7 (0.6)
	solution[d]	forward	6.6 + 11.4	397.4 (0.1)	2.6 (0.5)	399.9 (0.5)
	solution[d]	backward	6.6 + 11.4	396.8 (0.7)	2.6 (0.4)	399.4 (0.8)

Free energies in kcal/mol. Apparent statistical uncertainty in parentheses. Forward runs remove HPO_4^{2-}; backward runs introduce it; the free energies are for Pi deletion. Run lengths correspond to the sum of all windows for the electrostatic + van der Waals stages.

[a]Runs using a different procedure for the initial Pi placement; see text.

[b]In the presence an associated Mg^{2+} ion, from earlier work [10].

[c]Runs performed with a neutral His97; the other runs treat His97 as doubly-protonated (its most stable state; see Table 1).

[d]In the absence an associated Mg^{2+} ion, from earlier work [10].

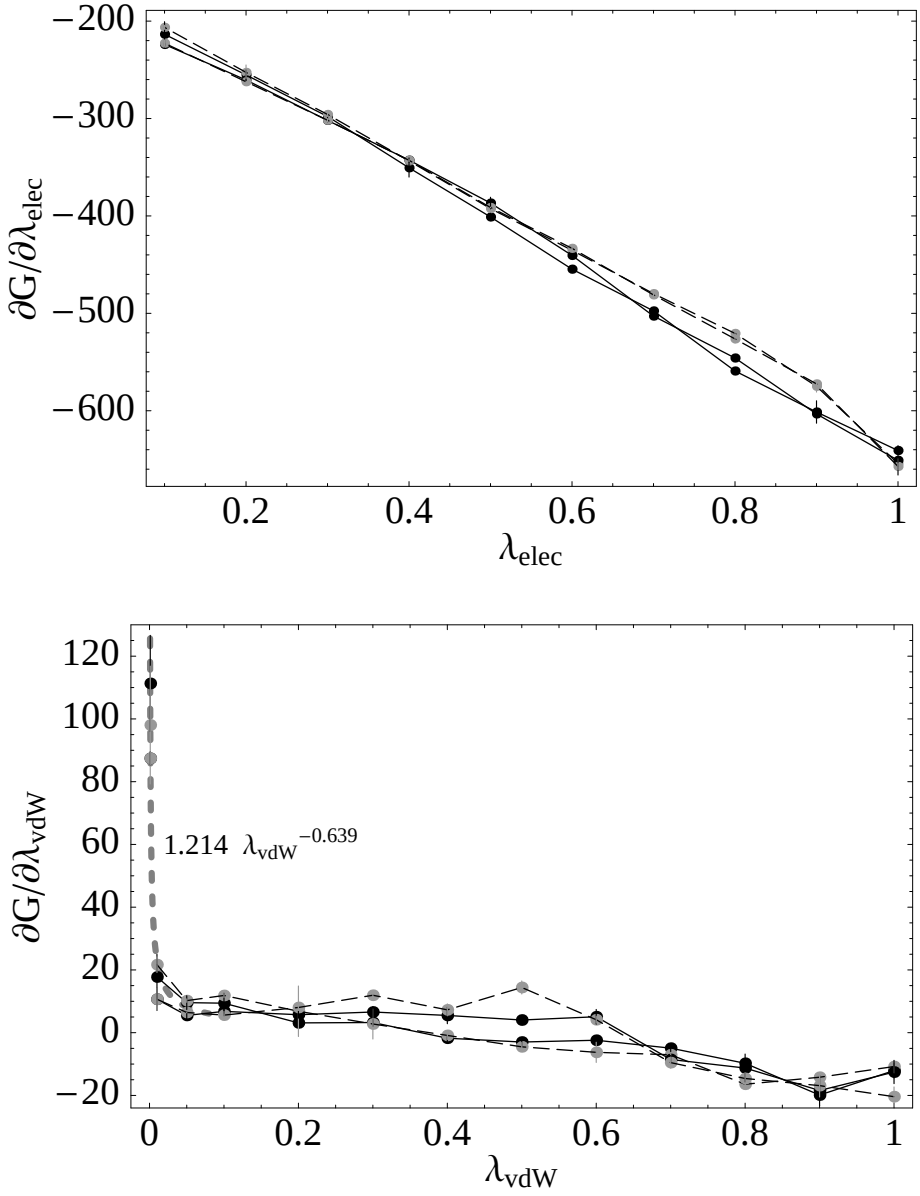

Figure 1: Free energy derivatives for the four MIXED state runs 9–12 (in kcal/mol). Backward runs are shown using grey dots and dashed lines. Error bars are shown (but frequently too small to be seen). The function used to fit the derivative close to $\lambda_{\mathrm{vdW}} = 0$ is shown as a thicker dashed line.

References

[1] Sham, Y., Chu, Z., and Warshel, A. Consistent calculations of pK_a's of ionizable residues in proteins: semi-microscopic and microscopic approaches. *J. Phys. Chem. B 101* (1997), 4458–4472.

[2] Eberini, I., Baptista, A., Gianazza, E., Fraternali, F., and Beringhelli, T. Reorganization in apo- and holo-β-lactoglobulin upon protonation of Glu89: molecular dynamics and pK_a calculations. *Proteins 54* (2004), 744–758.

[3] Archontis, G., and Simonson, T. Proton binding to proteins: a free energy component analysis using a dielectric continuum model. *Biophys. J. 88* (2005), 3888–3904.

[4] Aleksandrov, A., Proft, J., Hinrichs, W., and Simonson, T. Protonation patterns in tetracycline:Tet repressor recognition: Simulations and experiments. *ChemBioChem 8* (2007), 675–685.

[5] Im, W., Beglov, D., and Roux, B. Continuum solvation model: computation of electrostatic forces from numerical solutions to the Poisson-Boltzmann equation. *Comp. Phys. Comm. 111* (1998), 59–75.

[6] Simonson, T. Free energy of particle insertion. an exact analysis of the origin singularity for simple liquids. *Molec. Phys. 80* (1993), 441–447.

[7] Steinbrecher, T., Mobley, D. L., and Case, D. A. Nonlinear scaling schemes for Lennard-Jones interactions in free energy calculations. *J. Chem. Phys. 127* (2007), 214108.

[8] Bruckner, S., and Boresch, S. Efficiency of alchemical free energy simulations II: Improvements for thermodynamic integration. *J. Comp. Chem. 32* (2011), 1320–1333.

[9] Kastenholz, M. A., and Hünenberger, P. H. Computation of methodology-independent ionic solvation free energies from molecular simulations. II the hydration free energy of the sodium cation. *J. Chem. Phys. 124* (2006), 224501.

[10] Satpati, P., Clavaguéra, C., Ohanessian, G., and Simonson, T. Free energy simulations of a GTPase: GTP and GDP binding to archaeal Initiation Factor 2. *J. Phys. Chem. B 115* (2011), 6749–6763.