Supporting Information

Local Ordering at the N–H Sites of the Rho GTPase Binding Domain of plexin-B1: the Impact of Dimerization

Netanel Mendelman,1 Mirco Zerbetto,2 Matthias Buck3 and Eva Meirovitch1,*

The Mina and Everard Goodman Faculty of Life Sciences, Bar-Ilan University, Ramat-Gan 52900 Israel1; Department of Chemical Sciences, University of Padova, Padova 35131, Italy2; Case Western Reserve University. Department of Physiology and Biophysics, Cleveland OH, USA3

Corresponding author: eva.meirovitch@biu.ac.il, phone: 972-3-531-8049
Derivation of the real spherical harmonics, $Y_{L,K}$, from the complex spherical harmonics, $\mathbf{Y}_{L,K}$.

For $L = 1$ one has:

\[
Y_{1, -1} = p_y = i \frac{2}{\sqrt{2}} (Y_{1}^{-1} + Y_{1}^{1}) = i \frac{2}{\sqrt{2}} (2Imag Y_{1}^{1}) = \frac{\sqrt{3}}{4\pi} \cdot \frac{y}{r}
\]

\[
Y_{1,0} = p_z = Y_{1}^{0} = \frac{\sqrt{3}}{4\pi} \cdot \frac{z}{r}
\]

\[
Y_{1,1} = p_x = i \frac{1}{\sqrt{2}} (Y_{1}^{-1} - Y_{1}^{1}) \equiv i \frac{1}{\sqrt{2}} (2Re Y_{1}^{1}) = \frac{\sqrt{3}}{4\pi} \cdot \frac{x}{r}
\]

The functions $Y_{1}^{-1}, Y_{1}^{0}$ and $Y_{1}^{1}$ are the (complex) spherical harmonics for $L = 1$ and $K = -1, 0$ and 1; the functions $Y_{1, -1}, Y_{1,0}$ and $Y_{1,1}$ are the corresponding real combinations. The complex Wigner rotation matrix elements, $D_{0}^{1}_{-1}, D_{0}^{1}_{1},$ and $D_{0}^{1}_{01}$, are in direct proportion to the functions $Y_{1}^{-1}, Y_{1}^{0}$ and $Y_{1}^{1}$; the real Wigner functions $D_{1, -1}, D_{1,0}$ and $D_{1,1}$ are in direct proportion to the functions $Y_{1, -1}, Y_{1,0}$ and $Y_{1,1}$. In the Principal Axes System of the local ordering tensor only $D_{1,0}$ and $D_{1,1}$ survive.

For $L = 2$ one has:

\[
Y_{2, -2} = d_{xy} = i \frac{3}{\sqrt{2}} (Y_{2}^{-2} - Y_{2}^{2}) \equiv i \frac{3}{\sqrt{2}} (2Imag Y_{2}^{2}) = \frac{1}{2\sqrt{\pi}} \cdot \frac{3\sqrt{15}}{r^2} \cdot xy
\]

\[
Y_{2, -1} = d_{yz} = i \frac{2}{\sqrt{2}} (Y_{2}^{-1} + Y_{2}^{1}) \equiv i \frac{2}{\sqrt{2}} (2Imag Y_{2}^{1}) = \frac{1}{2\sqrt{\pi}} \cdot \frac{2\sqrt{15}}{r^2} \cdot yz
\]

\[
Y_{2, 0} = d_{z^2} = Y_{2}^{0} = \frac{1}{2} \frac{5}{\sqrt{4}} \cdot \frac{-x^2 - y^2 + 2z^2}{r^2}
\]

\[
Y_{2, 1} = d_{xz} = \frac{3}{\sqrt{2}} (Y_{2}^{-1} - Y_{2}^{1}) \equiv \frac{3}{\sqrt{2}} (2Re Y_{2}^{1}) = \frac{1}{2\sqrt{\pi}} \cdot \frac{3\sqrt{15}}{r^2} \cdot xz
\]

\[
Y_{2, 2} = d_{x^2 - y^2} = \frac{3}{\sqrt{2}} (Y_{2}^{-2} + Y_{2}^{2}) \equiv \frac{3}{\sqrt{2}} (2Re Y_{2}^{2}) = \frac{1}{4\sqrt{\pi}} \cdot \frac{3\sqrt{15}}{r^2} \cdot x^2 - y^2
\]

The functions $Y_{2}^{-2}, Y_{2}^{-1}, Y_{2}^{0} Y_{2}^{1}$ and $Y_{2}^{2}$ are the (complex) spherical harmonics for $L = 2$ and $K = -2, -1, 0, 1$ and 2; the functions $Y_{2, -2}, Y_{2, -1}, Y_{2, 0}, Y_{2, 1}$ and $Y_{2, 2}$ are the corresponding real combinations. The complex Wigner rotation matrix elements, $D_{0}^{2}_{-2}, D_{0}^{2}_{-1}, D_{0}^{2}_{00}, D_{0}^{2}_{01}$ and $D_{0}^{2}_{02}$
, are in direct proportion to the functions $Y_{2,-2}$, $Y_{2,-1}$, $Y_{2,0}$, $Y_{2,1}$ and $Y_{2,2}$; the real Wigner functions $D_{2,-2}$, $D_{2,-1}$, $D_{2,0}$, $D_{2,1}$ and $D_{2,2}$ are in direct proportion to the functions $Y_{2,-2}$, $Y_{2,-1}$, $Y_{2,0}$, $Y_{2,1}$ and $Y_{2,2}$. In the Principal Axes System of the local ordering tensor only $D_{2,0}$ and $D_{2,2}$ survive.

For $L = 3$ and $L = 4$ analogous situations are in effect.

**Table S1.** Character Table of the $D_{2h}$ point-group.¹

<table>
<thead>
<tr>
<th></th>
<th>E</th>
<th>C$_2$(z)</th>
<th>C$_2$(y)</th>
<th>C$_2$(x)</th>
<th>i</th>
<th>σ(xy)</th>
<th>σ(xz)</th>
<th>σ(yz)</th>
<th>Linear rotations</th>
<th>quadratic</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_g$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>$R_z$</td>
<td>$x^2, y^2, z^2$</td>
</tr>
<tr>
<td>$B_{1g}$</td>
<td>1</td>
<td>1</td>
<td>$-1$</td>
<td>$-1$</td>
<td>1</td>
<td>1</td>
<td>$-1$</td>
<td>$-1$</td>
<td>$R_z$</td>
<td>$xy$</td>
</tr>
<tr>
<td>$B_{2g}$</td>
<td>1</td>
<td>$-1$</td>
<td>1</td>
<td>$-1$</td>
<td>1</td>
<td>$-1$</td>
<td>1</td>
<td>$-1$</td>
<td>$R_y$</td>
<td>$xz$</td>
</tr>
<tr>
<td>$B_{3g}$</td>
<td>1</td>
<td>$-1$</td>
<td>$-1$</td>
<td>1</td>
<td>1</td>
<td>$-1$</td>
<td>$-1$</td>
<td>1</td>
<td>$R_z$</td>
<td>$yz$</td>
</tr>
<tr>
<td>$A_u$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>$-1$</td>
<td>$-1$</td>
<td>$-1$</td>
<td>$-1$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$B_{1u}$</td>
<td>1</td>
<td>1</td>
<td>$-1$</td>
<td>$-1$</td>
<td>$-1$</td>
<td>$-1$</td>
<td>1</td>
<td>1</td>
<td>$z$</td>
<td></td>
</tr>
<tr>
<td>$B_{2u}$</td>
<td>1</td>
<td>$-1$</td>
<td>1</td>
<td>$-1$</td>
<td>1</td>
<td>$-1$</td>
<td>1</td>
<td>$-1$</td>
<td>$y$</td>
<td></td>
</tr>
<tr>
<td>$B_{3u}$</td>
<td>1</td>
<td>$-1$</td>
<td>$-1$</td>
<td>1</td>
<td>$-1$</td>
<td>1</td>
<td>1</td>
<td>$-1$</td>
<td>$x$</td>
<td></td>
</tr>
</tbody>
</table>
Potential of Mean Force from MD. The POMF, $u(\theta, \phi)$, is related to the atomistic internal energy $u(\theta, \phi, x)$ (where $x$ denotes coordinates of the 3D protein structure other than the angles $\theta$ and $\phi$), in terms of the relation $P_{eq}(\theta, \phi) = (\int dx \exp(-u(\theta, \phi, x)))/Z = \exp(-u(\theta, \phi))/Z$ ($I$ – integration domain). Solving this integral is problematic. MD can provide an approximation to $u(\theta, \phi)$; this is $u^{(MD)}(\theta, \phi)$, defined in terms of the estimate for the Boltzmann distribution, $P_{eq}^{(MD)}(\theta, \phi) = \exp(-u^{(MD)}(\theta, \phi))/Z$. Using a relatively long unbiased simulations, one derives $P_{eq}^{(MD)}(\theta, \phi)$ from the MD trajectory.

The procedure used to do so was developed in previous work. Four independent MD trajectories, each 110 (55) ns in length for the monomer (dimer) have been calculated and utilized. For each snapshot of a given MD simulation, the PAS of the inertia tensor, PF, is calculated and superposed on the first snapshot, thus eliminating the global motion of the protein. Then $k$-clustering based on torsion-angle RMSD is performed, and the snapshot closest to the center of the most populated cluster is selected as reference structure. The latter is considered to represent the minimum free energy configuration.

The uniaxial local director frame, VF, is defined in the reference structure for every N–H site (details appear in the main text). For the monomer (dimer) the rotation matrix $R_{PF \rightarrow VF}$ is derived from the atomic coordinates of the trajectory with reference to this structure. The (fixed) Euler angles $\Omega_{OF \rightarrow DF}$, where DF is the $^{15}$N–$^1$H dipolar frame with $Z_{DF}$ along the N–H bond, are taken to be $\Omega_{OF \rightarrow DF} = (0, -101.3^\circ, -90^\circ)$.

In each snapshot one defines the OF and DF frames. The time-dependent matrix transforming PF into OF is extracted from the MD trajectory. Subsequently the time-series of the Euler angles $\Omega_{VF \rightarrow OF}(t)$ is derived by applying the appropriate Wigner rotations. The function $P_{eq}^{(MD)}(\theta, \phi)$ is obtained as histogram directly from the MD trajectory. Further details appear in ref 2.

Optimization of $P_{eq}^{(D_{ijk})}(\theta, \phi)$ versus $P_{eq}^{(MD)}(\theta, \phi)$. The MD-derived $P_{eq}^{(MD)}(\theta, \phi)$ functions consist of 360×360 squared pixels, to be converted into 360×360 numerical matrices. The matrix representations of
the $P_{eq}^{D_{L,K}}(\theta, \varphi)$ functions must have the same dimensions. Under these circumstances the optimization is quite compute-intensive. We found that one may reduce the matrix size to 46$\times$46 by dividing the original 2D $P_{eq}^{(MD)}(\theta, \varphi)$ pictures into equal areas, and replacing the latter by their centers. The results are virtually the same as those obtained using 360$\times$360 matrices, and the calculations are substantially faster (minutes instead of hours). The authors of ref 5, who treated $P_{eq}^{(MD)}(\theta, \varphi)$ functions in a different context, used 41$\times$41 matrices.

To develop software that converts the colored $P_{eq}^{(MD)}(\theta, \varphi)$ pictures into numerical matrices, we used the $P_{eq}^{(MD)}(\theta, \varphi)$ function of the N–H bond of Gln56 as test-case.\(^2\) The scheme developed features a color-ruler (normalization and the fact that probability distributions must be positive are properly addressed). The objective is to determine for each N–H bond the particular $u^{(D_{L,K})}(\theta, \varphi)$ function for which $P_{eq}^{(D_{L,K})}(\theta, \varphi) = \exp(-u^{(D_{L,K})}(\theta, \varphi))/Z$ “best fits” $P_{eq}^{(MD)}(\theta, \varphi)$. For that we used the “Sequential Quadratic Programming (SQP)” iterative method for constrained non-linear least-squares optimization from the “Global Optimum Solution MultiStart” MATLAB functionality.\(^6\) The expression:

$$\chi^2 = \sum_{(\theta, \varphi)} (P_{eq}^{(MD)}(\theta, \varphi)_i - P_{eq}^{(D_{L,K})}(\theta, \varphi)_i)^2$$

is minimized in this algorithm.

In order to initiate an optimization process, a “starting probability distribution”, $P_{eq}^{(D_{L,K})}$, i.e., a starting potential, $u^{(D_{L,K})}$, is required. Various forms of $u^{(D_{L,K})}$, i.e., various linear combinations of the $D_{L,K}$ functions, were considered. Terms $c_{L,K} D_{L,K}$ were assembled systematically in order of increasing $L$ and $K$. After a number of trials quite clear trends, guiding the search for appropriate starting potentials,
u(D_{L|K}), hence appropriate starting probability distributions, P_{eq}^{(D_{L|K})}, emerged. Although conducted manually, we estimate the search to have been virtually exhaustive.

The coefficients c_{L|K}, which are the weighting factors of the D_{L|K} functions in the linear combinations, u^{(D_{L|K})}, were allowed to vary in the (−30, +30) range. These limits were determined on the basis of previous findings showing that order parameters attain their extreme values when the potentials defining them level off in absolute value at roughly 30 kT. With the function P_{eq}^{(MD)} as “target” and the function P_{eq}^{(D_{L|K})} as variable element, we carried out global minimization. Once the “best-fit” function, P_{eq}^{(D_{L|K}−BEST)}, was determined, a colored picture thereof was created and compared with the corresponding MD-derived function, P_{eq}^{(MD)}.

Goodness-of-fit. We devised four criteria for evaluating the “goodness of the fit”: χ², R-squared (value of 1 for perfect fit), visual agreement, and difference map given by P_{eq}^{(MD)} − P_{eq}^{(u(BEST))}. Good visual agreement and close to zero (i.e., totally blue area) difference maps were obtained for χ² predominantly smaller than 2.5 and R-squared larger than 0.97.

The MD-derived P_{eq}^{(MD)}(θ,φ) functions are obtained on the basis of statistical analysis of the MD trajectory. Necessarily the P_{eq}^{(MD)}(θ,φ) pictures comprise areas where the intensity should be zero. However, in actual fact, quite a few points in these regions emerge from the MD simulations with small but finite intensity, causing dependence of χ² on the magnitude of these regions. By nearly equalizing the latter, meaningful relative χ² values, and a χ² threshold of typically 2.5 for acceptance, have been determined.

Error margin. Based on the residual dependence of the coefficients c_{L|K} on the magnitude of the blue areas, and changes in “best-fit” c_{L|K} values which still preserve the visual agreement between P_{eq}^{(D_{L|K}−BEST)} and P_{eq}^{(MD)}(θ,φ), we estimate the error in the “best-fit” c_{L|K} parameters at 10%.
**Figure S1.** $P_{eq}^{(MD)}$ histograms for the β₁-strand of the plexin-B1 RBD in monomer and dimer form.

**Figure S2.** $P_{eq}^{(MD)}$ histograms for residues 14–19 of the L1-loop of the plexin-B1 RBD in monomer and dimer form.
**Figure S3.** $P_{eq}^{(MD)}$ histograms for residues 20–26 of the L1-loop of the plexin-B1 RBD in monomer and dimer form.

**Figure S4.** $P_{eq}^{(MD)}$ histograms for the $\beta_2$-strand of the plexin-B1 RBD in monomer and dimer form.
**Figure S5.** $P^{(MD)}_{\text{eq}}$ histograms for the $\beta_2/\alpha_1$-turn of the plexin-B1 RBD in monomer and dimer form.

**Figure S6.** $P^{(MD)}_{\text{eq}}$ histograms for residues 36–41 of the $\alpha_1$-helix of the plexin-B1 RBD in monomer and dimer form.
Figure S7. $P^{(MD)}_{eq}$ histograms for residues 42–47 the $\alpha_1$-helix of the plexin-B1 RBD in monomer and dimer form.

Figure S8. $P^{(MD)}_{eq}$ histograms for residues 48–55 of the L2-loop of the plexin-B1 RBD in monomer and dimer form.
Figure S9. $P_{eq}^{(MD)}$ histograms for residues 56–63 of the L2-loop of the plexin-B1 RBD in monomer and dimer form.

Figure S10. $P_{eq}^{(MD)}$ histograms for the $\beta_3$-strand of the plexin-B1 RBD in monomer and dimer form.
**Figure S11.** $P_{eq}^{(MD)}$ histograms for the L3-loop of the plexin-B1 RBD in monomer and dimer form.

**Figure S12.** $P_{eq}^{(MD)}$ histograms for the $\beta_4$-strand of the plexin-B1 RBD in monomer and dimer form.
Figure S13. $P_{eq}^{(MD)}$ histograms for residues 78–83 of the L4-loop of the plexin-B1 RBD in monomer and dimer form.

Figure S14. $P_{eq}^{(MD)}$ histograms for residues 84–89 of the L4-loop of the plexin-B1 RBD in monomer and dimer form.
Figure S15. $P_{eq}^{(MD)}$ histograms for residues 90–95 of the L4-loop of the plexin-B1 RBD in monomer and dimer form.

Figure S16. $P_{eq}^{(MD)}$ histograms for the $\alpha_2$-helix of the plexin-B1 RBD in monomer and dimer form.
Figure S17. $P_{eq}^{(MD)}$ histograms for the $\alpha_2/\beta_5$-turn of the plexin-B1 RBD in monomer and dimer form.

Figure S18. $P_{eq}^{(MD)}$ histograms for the $\beta_5$-strand of the plexin-B1 RBD in monomer and dimer form.

The highly flexible N-, and C-terminal segments have been disregarded.
References

1. *e-Chemical Portal, D_{2h} point group,* [http://www.webqc.org/printable-symmetrypointgroup-d2h.html](http://www.webqc.org/printable-symmetrypointgroup-d2h.html).


