Protocol for modeling the Ube2g2:Ub and Ube2g2:K48-diUb interactions using Rosetta

Docking of ubiquitin and diubiquitin to Ube2g2 was carried out using a version of the RosettaDock protocol (1) in which the score functions were modified to award a bonus to models satisfying observed CSPs and a penalty for contacts that were not reflected by CSPs (unobserved CSPs). To incorporate the CSP constraints, the interfacial residues (defined as residues belonging to different chains with C_α atoms within 8 Å of each other) were detected, and intersecting residues of observed (unobserved) CSP were counted. The distance of 8 Å was chosen as an approximation of side chain length based on visual inspection of the solution structure of the UbcH5c/Ub non-covalent complex (2) (PDB ID: 2FUH), and was not varied between residues of different types. The number of observed (unobserved) CSPs found in the interfacial residues was divided by the total number of observed (unobserved) CSPs to obtain the fractional observed (unobserved) CSP. The fraction observed (unobserved) was multiplied by five and subtracted (added) from the Rosetta low resolution score (in low-resolution, all atoms of the side chain of each residue were represented by a pseudo-atom known as the centroid) to provide a bonus (penalty) in the low resolution stage. For the Rosetta full-atom score in high resolution, the fraction was multiplied by one-thousand. A model was considered to satisfy PRE constraints if all residue pairs detected to be in proximity by PRE experiments had a C_α-C_α distance within 15 Å, otherwise the model was penalized by adding a weighted bounded constraint function to its score for each detected residue pair. The constraint function is defined as follows:

$$f(x)=\begin{cases}\left(\dfrac{x-x_{\min}}{\sigma}\right)^2 & \text{for } x < x_{\min} \\ 0 & \text{for } x_{\min} \le x \le x_{\max} \\ \left(\dfrac{x-x_{\max}}{\sigma}\right)^2 & \text{for } x_{\max} < x \le \left(x_{\max}+r_{switch}\sigma\right) \\ \dfrac{1}{\sigma}\left(x-\left(x_{\max}+r_{switch}\cdot\sigma\right)\right)+\left(\dfrac{\sigma\cdot r_{switch}}{\sigma}\right)^2 & \text{for } x > \left(x_{\max}+r_{switch}\sigma\right)\end{cases} \quad (1)$$

Where x is the distance between the C_α atoms of a detected residue pair, x_{max} is the maximum value permitted before the penalty is incurred (15 Å), x_{min} is the minimum value permitted before the penalty is incurred, σ is the standard error of the PRE distances (4 Å), and r_{switch} is a constant (= 0.5).

For the diubiquitin simulation the same function was used to constrain the K48 isopeptide bond based on the C_α-C_α distance between lysine 48 of the proximal ubiquitin and glycine 76 of the distal ubiquitin. For the isopeptide bond, x_{min} is set to 0 Å, x_{max} to 10 Å and σ to 2 Å. The constraint function were weighted by 10 and 100 for the low-resolution and full-atom scoring functions, respectively.

The detailed protocol employed for the ubiquitin and K48-linked diubiquitin models is as follows:

i. Starting Configuration: The simulations were started with the ubiquitin oriented as found in the PDB structure 2FUH. For diubiquitin simulations, the distal ubiquitin monomer was placed in the 2FUH orientation, and the proximal monomer was manually oriented by visual satisfaction of the PRE data.

ii. Starting Perturbation:

   a. Ubiquitin: (1) Ubiquitin was perturbed by a rigid-body move (Gaussian perturbations of standard deviation of 5 Å translation and 15° rotation) along with a rotational degree of freedom (spin) along the axis connecting the center of mass of the enzyme and the ubiquitin.

Similar to RosettaDock, the partners were moved towards each other to make glancing contact. The conformation was perturbed using fifty standard centroid-mode RosettaDock rigid-body moves (Gaussian perturbation of ~1 Å translation and ~10° rotation) using the constrained score (Eq. 6) to evaluate each conformer during the Monte Carlo acceptance criterion. (2) The rigid-body perturbation was followed by perturbation of the four C-terminal residues. The flexible residues were first initialized with ideal bond lengths and angles and an extended conformation. The C-terminal backbone torsional angles were perturbed by forty cycles of eight "small" and eight "shear" moves(3) (Gaussian 90° perturbation) each to account for the large flexibility of this region of the ubiquitin. The conformation after each cycle was accepted based on the Metropolis criterion, and lowest scoring conformation observed during the forty cycles is chosen as the final conformation. During this first stage of C-terminal perturbation, the rigid-body orientations of the ubiquitin and the enzyme were kept fixed.

b. Diubiquitin: The four C-terminal residues were removed from both ubiquitin monomers. First, only one of the monomers is considered. If the distal monomer was selected, it was perturbed by a conservative rigid-body perturbation (Gaussian perturbation of 1 Å translation and 3° rotation), and if the proximal monomer was selected, it was perturbed by larger rigid-body perturbations (5 Å translation and 15° rotation) with a rotational degree of freedom (spin) along the axis connecting the center of mass of the enzyme and the proximal ubiquitin. The Ub and Ube2g2 partners were moved into glancing contact. The selected monomer was perturbed using fifty standard centroid-mode RosettaDock rigid-body moves (1 Å translation and 10° rotation) using the unconstrained low-resolution score to evaluate each conformer during the Monte Carlo acceptance criterion. The lowest-scoring orientation observed during the course of the fifty cycles is restored, and then the other ubiquitin monomer (without the C-terminal tail) is brought back into the system in the starting orientation and moved with the same set of rules as above. If the second ubiquitin monomer is the proximal subunit, the spin is applied along the axis connecting the center of mass of the enzyme-distal ubiquitin complex and the proximal ubiquitin. The originally selected monomer was perturbed using fifty standard centroid-mode RosettaDock rigid-body moves (Gaussian perturbation of ~1 Å translation and ~10° rotation) using the unconstrained low-resolution score (which accounts for all inter-chain interfaces) with the CSP (non-CSP) bonus (penalty) to evaluate each conformer during the Monte Carlo acceptance step. Then the C-terminal residues are restored in their original conformation. The rigid-body perturbation was followed by perturbation to the conformation of the four C-terminal residues as described in (ii.a.2). If the distance between the C_α atoms of lysine 48 of the proximal ubiquitin and glycine 76 of the distal ubiquitin is more than 1.25 x 15.0 Å (18.75 Å), the structure is rejected, and the algorithm is restarted.

iii. Low resolution search:

a. Ubiquitin: Fifty cycles of randomly selected moves consisting of either (1) rigid-body (Gaussian 0.7 Å translation and 5° rotational; 75% of moves) or (2) C-terminal perturbation (25% of moves) sampled the conformational space in the immediate vicinity of the perturbed starting conformation. The C-terminal perturbation incorporated ten perturbation cycles followed by a Metropolis test based on the constrained low-resolution scoring function. Each perturbation cycle comprised five rounds of small and shear moves (Gaussian 90° perturbation), followed by a line minimization in the backbone coordinates of the four C-terminal residues.

b. Diubiquitin: Fifty cycles of randomly selected moves consisting of either: (1) distal (37.5% of moves) or (2) proximal (37.5%) subunit rigid-body (0.7 Å translation and 5° rotational) moves or (3) C-terminal perturbation (as in iii.a, 25%) were used to sample exhaustively the

conformational space in the immediate vicinity of the perturbed starting conformation.

iv. Filters: The structure with the lowest constrained score observed during the fifty cycles was selected for full-atom refinement only if it had (1) a negative (unconstrained) low-resolution score, (2) satisfied at least 70% of the observed CSP, and (3) had a PRE restraint score of less than 4.5(1.5) based on equation 6 for mono(di)-ubiquitin simulations. The ubiquitin complex had thrice the number of PRE restraints (~13) as compared to the diubiquitin complex (~4), accounted for by tripling the cutoff for the ubiquitin PRE restraint score. For diubiquitin simulations, an additional filter (4) is imposed to reject models that do not have the K48 isopeptide bond: a model is rejected if the C_α distance between lysine 48 of the proximal ubiquitin and glycine 76 of the distal ubiquitin is more than 15.0 Å. If a model failed any filters, the model was discarded and the algorithm was restarted.

v. The low-resolution centroid-mode model was changed to a high-resolution full-atom model by copying the side-chain conformations from the input structure and repacking rotamers at the inter-chain interface and in and around the four C-terminal residues. Rotamer packing is a Monte Carlo simulated annealing(4) scheme to select the most favorable side chain conformations from a discrete library.(5)

vi. High resolution search:

   a. Ubiquitin: The full-atom model was optimized by 100 cycles of randomly selected rigid-body moves (75% probability, 0.1 Å translation and 5° rotation) and C-terminal perturbation (25% probability, small and shear moves of 5° magnitude) with Davidon-Fletcher-Powell minimization(6) of the terminus and nearby atoms). Each move is followed by side-chain optimization by rotamer trials or full repacking every eight cycles.(7)

   b. Diubiquitin: (1) The C-terminal tails of both ubiquitin monomers are optimized by ten cycles of perturbation (as in vi.a), repacking, and a Metropolis Monte Carlo test after each cycle using the constrained full-atom scoring function. The full-atom model was further optimized by 100 cycles of randomly selected (2) distal or (3) proximal rigid-body moves (37.5% probability each) and (4) C-terminal perturbation (25% probability) with side chain optimization by rotamer trials or full repacking every eight cycles.

vii. Final optimization: The model with the lowest full-atom score (including the CSP/non-CSP bonus/penalty) at the end of the 100 cycles was subjected to rotamer packing of the interfacial and C-terminal and neighboring residues and subsequent optimization by sampling off-rotamer side-chain conformations by torsion space minimization.(7)

viii. High-resolution filtering: The model was retained only if it satisfied the filters described in (iv), now using all-atom scores. Additionally, the unconstrained interface energy (score of the complex minus the score of the isolated monomers) of each model was used as an additional metric.

ix. One thousand models are generated and sorted by the unconstrained Rosetta full-atom scoring function. The largest cluster of similar models amongst the ten lowest-scoring models is chosen as the solution set. A representative model is selected from the cluster based on a negative interface energy and the lowest value of the constraint score.

x. Linker rebuilding: For the diubiquitin docking simulations, the K48 isopeptide bond and the match with PRE observations for the MTSL tag attached to the 75^{th} residue of the proximal subunit is optimized in a post-processing simulation where the four C-terminal residues of each ubiquitin subunit is perturbed while keeping the backbone conformations of all other residues fixed. The constraint function defined earlier (Eq. 6) is used for optimization with the same parameters for the PRE constraints. For the K48 isopeptide bond, the C_α-C_α distance has an x_{min},

x_{max} and σ of 5 Å, 10 Å and 2 Å, respectively. A weighted harmonic constraint is used on the distance between the N_ζ of K48 in the proximal Ub and the O at the terminal G76 of the distal Ub as:

$$f(x) = ((x - x_0) / \sigma)^2 \quad (2)$$

xi. Where x_0 is the average bond distance (2.7 Å) and σ is the standard deviation (2.0 Å). The weight of the harmonic potential is set to 1,000,000. The C-terminus was optimized by 100 cycles of perturbations and repacking as described in (vi.b.1). The model with the lowest constrained full-atom score (with CSP/non-CSP bonus/penalty) is output if it satisfies filters specified in (viii); otherwise it is rejected and the algorithm is restarted from step (i). The final model is selected based on criteria in (ix).

References:

1. Gray, J. J., Moughon, S., Wang, C., Schueler-Furman, O., Kuhlman, B., Rohl, C. A., and Baker, D. (2003) *J. Mol. Biol.* **331**, 281-299
2. Brzovic, P. S., Lissounov, A., Christensen, D. E., Hoyt, D. W., and Klevit, R. E. (2006) *Mol. Cell* **21**, 873-880
3. Rohl, C. A., Strauss, C. E. M., Misura, K. M. S., Baker, D., Ludwig, B., and Michael, L. J. (2004) Protein Structure Prediction Using Rosetta. in *Numerical Computer Methods, Part D*, Academic Press
4. Kuhlman, B., and Baker, D. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 10383-10388
5. Dunbrack, R. L., and Karplus, M. (1993) *J. Mol. Biol.* **230**, 543-574
6. Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1992) *Numerical Recipes in C*, Cambridge University Press, Cambridge
7. Wang, C., Schueler-Furman, O., and Baker, D. (2005) *Protein Sci.* **14**, 1328-1339