QASDOM meta-server for automatic analysis, scoring and ranking of docking models.

Availability: http://qasdom.eimb.ru

Ranking scores

We define that an intermolecular atomic contact occurs when the distance between two atoms is less then the cut-off value. By default the value is 4.5 Å. Plausible values range from 4 to 6 Å. A residue contact is established when at least one atom in the residue forms atomic contact.

If the dataset consists of ***M*** models, where ***m*** is an individual model number, and further if ***L*** is the length of the amino acid sequence in the models where ***l*** is the number of a residue (sequence position) in this sequence, then the number of atomic contacts of the amino acid residue ***l*** in the model ***m*** is AC_{ml}. Formally, residue contact of the amino acid residue ***l*** in the model ***m*** is determined by the following condition

$$C_{ml} = \begin{cases} 1, \mathrm{AC}_{ml} > 0 \\ 0, \mathrm{AC}_{ml} = 0 \end{cases}$$

The number of residue contacts for the amino acid ***l*** in all models in the dataset is defined as

$$RC_l^{aa} = \sum_{m=1}^{M} C_{ml}$$

The value of RC_l^{aa} is calculated as a number of models in which the amino acid residue ***l*** participates in an interaction.

Accordingly, the number of residue contacts (amino acid residues participating in interactions) in the model ***m*** is

$$RC_m^{\mathrm{mod}} = \sum_{l=1}^{L} C_{ml}$$

The total number of receptor-ligand amino acid residue interactions in the dataset of models is

$$C_{total} = \sum_{l=1}^{L} \sum_{m=1}^{M} C_{ml}$$

We introduce a score for each model ***m***, which reflects the degree of representativeness of this model for the overall dataset of models. We assign to each amino acid residue ***l*** involved in interaction in the model ***m***, the number of models in which the amino acid residue ***l*** participates in the interaction. These values are summed up along the entire sequence ***L*** and normalised by the total number of amino acid residues interacting with the ligand in the entire dataset of models

$$S_m = \frac{1}{C_{total}} \sum_{l=1}^{L} C_{ml} \cdot RC_l^{aa} \qquad (1)$$

In terms of residue contacts, S_m score reflects the degree of representativeness of the interacting residues of the model ***m*** in this dataset of models, i.e. similarity of the model ***m*** to a consensus model built for this dataset of models.

S_m values can range from $1/C_{total}$ in the case where there is only one interacting amino acid in the model that is not observed in other models, up to **1** when all models in that dataset are identical and form identical interactions. If there is no contacts between amino acid residues of the receptor and ligand in the model ***m***, S_m for the model is assumed to be zero and the model ***m*** does not participate in calculating the scores for other models.

In addition to the score S_m (1) we introduce a similar score S_m^{atomic}, based on the number of atomic contacts AC_{ml} of amino acid residues in the model m

$$S_m^{atomic} = \frac{1}{C_{total}^{atomic}} \sum_{l=1}^{L} (AC_{ml} \cdot \sum_{m=1}^{M} C_{ml}) \qquad (2)$$

This score can have values above 1 for the models where the number of atomic contacts is higher than the mean for the dataset, and the sets of interacting residues are close to the consensus model. The S_m^{atomic} score can be either used independently when ranging on the atomic contacts is preferable, or to fine-tune the S_m score. I.e. in the case of equal S_m score values one can use the S_m^{atomic} score to refine ranking and choose the model with a higher S_m^{atomic} value. The S_m^{atomic} score provides possibilities for a more detailed comparison of models.

In the following example the models are identical, with the same atomic interactions observed in the same amino acid residues. Then for the dataset of M identical models

$$S_m^{atomic} = \frac{1}{C_{total}^{atomic}} \sum_{l=1}^{L} (AC_{ml} \cdot \sum_{m=1}^{M} C_{ml}) = \frac{M}{C_{total}^{atomic}} \sum_{l=1}^{L} AC_{ml} = \frac{M \cdot RC_m^{atomic}}{C_{total}^{atomic}} = 1$$

for all models. If one atomic contact is added to the model n from the M dataset while the number of amino acid residue contacts remains the same, the S_m^{atomic} for model n will be

$$S_n^{atomic} = \frac{1}{C_{total}^{atomic} + 1} \sum_{l=1}^{L} (AC_{nl} \cdot \sum_{m=1}^{M} C_{ml}) = \frac{M \cdot (RC_m^{atomic} + 1)}{C_{total}^{atomic} + 1}$$

And for the identical models $m=1,\ldots, M, m \neq n$

$$S_m^{atomic} = \frac{1}{C_{total}^{atomic} + 1} \sum_{l=1}^{L} (AC_{ml} \cdot \sum_{m=1}^{M} C_{ml}) = \frac{M \cdot RC_m^{atomic}}{C_{total}^{atomic} + 1}$$

One can see that the S_m^{atomic} score is lower for all identical models compared to the increased score for the model n with one additional atomic contact.

Clustering procedure

The server automatically annotates two types of clusters, linear and structural.

Linear clusters are annotated in the sequences as consecutive groups of residues, using residue and atomic contacts. A cluster is formed (1) by the amino acids that participate in contacts with frequency above the *median* value for the relevant contacts in the dataset; (2) there can be no more then three consecutive gaps formed by residues within the cluster where frequency is below the median.

Structural cluster includes docking models with close interaction patterns that are defined by a distance matrix constructed for all models using residue interactions of each model. Elements of the matrix DM_{ij} are calculated as difference between the sum of the numbers of interacting residues in the models i and j, and the doubled number of common interacting residues in a pair of models ij defining this element.

$$DM_{ij} = RC_i^{\mathrm{mod}} + RC_j^{\mathrm{mod}} - 2 \cdot RC_i^{\mathrm{mod}} \bigcap RC_j^{\mathrm{mod}}$$

Thus, this measure is the sum of the number of non-overlapping interacting residues in the two compared models. The procedure begins with ranking of the DM_{ij} values. Clustering starts with models that have the smallest distance values. A model is added to the cluster if the distance between at least one model in the cluster and this model is below the median value calculated for the distance matrix. The clustering continues until $DM_{ij} < Median\{\underset{i,j=1}{\overset{M}{DM}}_{ij}\}$. Accordingly, some of the models remain in the single-model clusters.