READFA Documentation

1. Overview

Sampling of trajectory gives large amounts of pairwise force data with unignorable demands on memory and storage cost. For efficient storage and higher processing speed, forces data including pairwise forces with or without detailed components from potential terms and summed forces on atoms or residues are designed to write into binary serialized files. A program named *readFA* is developed for analyzing these serialized data with less memory usage and high processing efficiency. The program *readFA* also have functions of rotation correction based on conformational orientations, data format conversion and statistical analysis systematically.

Since the pairwise force data is generated with complex analysis group pairs configuration, these data are accumulated from pairwise forces between atom pairs on different levels of force units. The definition of pairwise forces is not restricted to forces between atom pairs or residue pairs, but is extended to that between a pair of force units

$$\mathbf{F}_{fui,fuj} = \sum_{i \in fui, j \in fuj} \mathbf{F}_{i,j}$$

where *i* and *j* are the atoms included in the force units *fui* and *fuj* respectively. It is worth mentioning that the two force units involved can be different. To avoid confusion in detailed analysis, topology information of how force unit is organized in simulation system is neglected. Instead, a unifying treatment of atoms, residues, groups, and the whole system is applied, and the virtual concept of force units is used throughout the organization of program *readFA*.

1.1 Rotation Correction Based on Conformational Orientations

Due to Brownian motion of solvent molecules, random collisions between water molecules and the protein molecule causing slow rotation of protein orientation during the sampling process. The summed forces on or pairwise forces between atoms or residues on protein are dependent on the coordinate system of protein orientations, while all force data is sampled under the Cartesian coordinate system of simulation box. The differences of coordinate systems in between result in difficult in analyzing force data as vectors, especially when the protein orientation based coordinate system is changing constantly throughout the sampled trajectory. By fitting the conformations of selected reference system, e.g., the protein itself, back to its initial conformation, vectors sampled under the the Cartesian coordinate system can be converted into protein orientation based coordinate system with a rotation operation. For convenience, the reference system at each frame is aligned to its conformation at *t* = *0*, or the first recorded conformation in the given trajectory, and a series of rotation matrix $\mathbf{U}_t$ are derived by quaternion–based 3D rotation algorithm. The $\mathbf{U}_t$ is decided by minimizing RMSD between the aligned structure $\widetilde{\mathbf{R}}_t = \mathbf{U}_t\mathbf{R}_t + \mathbf{s}_t$ and initial structure $\mathbf{R}_0$, where $\mathbf{R}_t$ refers to the reference structure at time *t*, and $\mathbf{s}_t$ represents the translation shift vector of reference system. The rotation matrix $\mathbf{U}_t$ is used to derive the converted force vector $\widetilde{\mathbf{F}_t} = \mathbf{U}_t\mathbf{F}_t$.

In practice, *readFA* program could perform rotation correction based on conformational orientations of given reference system. The protein molecule is usually selected as the reference system when studying the equilibrium state of a protein. More often, the Center of Mass (COM) data of each residue are used for determining the orientations of protein. Using all atoms data are less efficient for higher computational cost and limitations on catching the backbone motions.

1.2 Data Format Conversion

Force data derived from GROMACS simulations are in the units of kJ/mol/nm. *readFA* program converts the force data units to pN for convenience. The conversion between the two force units is 1 kJ/mol/nm ≈ 1.6605 pN. In addition, *readFA* program efficiently pack force and coordinate data on each force units (Table 1) or force unit pairs (Table 2) into standard CSV format text files.

Table 1 Detailed format of raw force and coordinate CSV files by *readFA* program on each force units.

Data Column	Expression	Description
frame		Conformation index starting from 0 in its trajectory order
i		Force unit index starting from 0
fx	$F_{i,x} = \sum_{j}^{N} F_{ij,x}$	Force X-component on force unit *i*
fy	$F_{i,y} = \sum_{j}^{N} F_{ij,y}$	Force Y-component on force unit *i*
fz	$F_{i,z} = \sum_{j}^{N} F_{ij,z}$	Force Z-component on force unit *i*

f	$\|\mathbf{F}_i\| = \sqrt{F_{i,x}^2 + F_{i,y}^2 + F_{i,z}^2}$	Force scalar on force unit *i*
x *	$R_{i,x}$	Cartesian coordinate X-component for force unit *i*
y *	$R_{i,y}$	Cartesian coordinate Y-component for force unit *i*
z *	$R_{i,z}$	Cartesian coordinate Z-component for force unit *i*

* Input of coordinates data are required.

Table 2 Detailed format of raw force and coordinate CSV files by *readFA* program on each force unit pairs.

Data Column	Expression	Description
frame		Conformation index starting from 0 in its trajectory order
i, j		Force unit index starting from 0
fx	$F_{ij,x}$	Force X-component on force unit pair *ij*
fy	$F_{ij,y}$	Force Y-component on force unit pair *ij*
fz	$F_{ij,z}$	Force Z-component on force unit pair *ij*
f	$\|\mathbf{F}_{ij}\| = \sqrt{F_{i,x}^2 + F_{i,y}^2 + F_{i,z}^2}$	Force scalar on force unit pair *ij*
type		Encoding force type of force unit pair *ij*
xi, xj *	$R_{i,x}/R_{j,x}$	Cartesian coordinate X-component for force unit pair *ij*
yi, yj *	$R_{i,y}/R_{j,y}$	Cartesian coordinate Y-component for force unit pair *ij*
zi, zj *	$R_{i,z}/R_{j,z}$	Cartesian coordinate Z-component for force unit pair *ij*

* Input of coordinates data are required.

1.3 Statistical Analysis

Number of forces collected are varied for different scenarios. We denote the number of force units as *N*, the number of pairwise forces between force units as *P* (extremely small pairwise forces are filtered out), and the number of correlation pairs between each force unit as *C(=N^2)*, respectively. A comparison of these numbers usually corresponds to the expression of $C \gg P > N$. In consideration of the number relationship, it is appropriate to separate calculations of these forces into different modules in program design. Therefore, *readFA* program performs calculations with different modules, with results written in different files. In total, *readFA* program is designed with three primary modules and several submodules. These three primary modules include summed force analysis module, pairwise force analysis module and correlation analysis module. There are also submodules include network (Graph) analysis module, regression module and sequence analysis module, and these submodules are still in development.

1.3.1 Summed Force Analysis Module

The summed force analysis module provides statistical analysis data of the forces $\mathbf{F}_i$ and positions $\mathbf{R}_i$ on each force unit, as well as correlation analysis between the forces $\mathbf{F}_i$ and the position offsets $\Delta\mathbf{R}_i$ (Table 3).

Table 1 Detailed format of summed force analysis results CSV files by *readFA* program on each force units.

<table>
<tr><th>Data Column</th><th>Expression</th><th>Description</th></tr>
<tr><td>i</td><td></td><td>Force unit index starting from 0</td></tr>
<tr><td><Fix> *</td><td>$\langle F_{i,x}\rangle$</td><td>Mean value of force X-component on force unit i</td></tr>
<tr><td><Fiy> *</td><td>$\langle F_{i,y}\rangle$</td><td>Mean value of force Y-component on force unit i</td></tr>
<tr><td><Fiz> *</td><td>$\langle F_{i,z}\rangle$</td><td>Mean value of force Z-component on force unit i</td></tr>
<tr><td><|Fi|> *</td><td>$\langle \lVert \mathbf{F}_i \rVert \rangle$</td><td>Mean value of force scalar on force unit i</td></tr>
<tr><td><|Fi|^2>^1/2 *</td><td>$\sqrt{\langle \lVert \mathbf{F}_i \rVert^2 \rangle} = \sqrt{\langle F_{i,x}^2 + F_{i,y}^2 + F_{i,z}^2 \rangle}$</td><td>Square root of mean value of force vectors norm squared on force unit i</td></tr>
<tr><td>|<Fi>| *</td><td>$\lVert \langle \mathbf{F}_i \rangle \rVert = \sqrt{\langle F_{i,x}\rangle + \langle F_{i,y}\rangle + \langle F_{i,z}\rangle}$</td><td>Scalar of mean force vector on force unit i</td></tr>
<tr><td>std(Fix) *‡</td><td>$\sigma_{F_{i,x}} = \sqrt{\frac{\sum_i^N \left(F_{i,x} - \langle F_{i,x}\rangle\right)^2}{N-1}}$</td><td>Sample standard deviation of force X-component on force unit i</td></tr>
<tr><td>std(Fiy) *‡</td><td>$\sigma_{F_{i,y}} = \sqrt{\frac{\sum_i^N \left(F_{i,y} - \langle F_{i,y}\rangle\right)^2}{N-1}}$</td><td>Sample standard deviation of force Y-component on force unit i</td></tr>
<tr><td>std(Fiz) *‡</td><td>$\sigma_{F_{i,z}} = \sqrt{\frac{\sum_i^N \left(F_{i,z} - \langle F_{i,z}\rangle\right)^2}{N-1}}$</td><td>Sample standard deviation of force Z-component on force unit i</td></tr>
<tr><td>std(|Fi|) *‡</td><td>$\sigma_{\lVert \mathbf{F}_i \rVert} = \sqrt{\frac{\sum_i^N \left(\lVert \mathbf{F}_i \rVert - \langle \lVert \mathbf{F}_i \rVert \rangle\right)^2}{N-1}}$</td><td>Sample standard deviation of force scalar on force unit i</td></tr>
<tr><td>|std(Fi)| *‡</td><td>$\lVert \sigma_{\mathbf{F}_i} \rVert = \sqrt{\sigma_{F_{i,x}}^2 + \sigma_{F_{i,y}}^2 + \sigma_{F_{i,z}}^2}$</td><td>—</td></tr>
<tr><td><DFi> *†</td><td>$\Delta F_i = \langle \lVert \mathbf{F}_i - \langle \mathbf{F}_i \rangle \rVert \rangle$</td><td>Mean value of force fluctuation magnitude on force unit i</td></tr>
<tr><td>std(DFi) *†‡</td><td>$\sigma_{\Delta F_i} = \sqrt{\frac{\sum_i^N \left(\Delta F_i - \langle \Delta F_i \rangle\right)^2}{N-1}}$</td><td>Sample standard deviation of force fluctuation magnitude on force unit i</td></tr>
</table>

<Rix> ⁑	$\langle R_{i,x} \rangle$	Average position X-coordinate for force unit *i*
<Riy> ⁑	$\langle R_{i,y} \rangle$	Average position Y-coordinate for force unit *i*
<Riz> ⁑	$\langle R_{i,z} \rangle$	Average position Z-coordinate for force unit *i*
std(Rix) ⁑‡	$\sigma_{R_{i,x}} = \sqrt{\frac{\sum_i^N \left(R_{i,x} - \langle R_{i,x} \rangle\right)^2}{N-1}}$	Position fluctuation in X-direction for force unit *i*
std(Riy) ⁑‡	$\sigma_{R_{i,y}} = \sqrt{\frac{\sum_i^N \left(R_{i,y} - \langle R_{i,y} \rangle\right)^2}{N-1}}$	Position fluctuation in Y-direction for force unit *i*
std(Riz) ⁑‡	$\sigma_{R_{i,z}} = \sqrt{\frac{\sum_i^N \left(R_{i,z} - \langle R_{i,z} \rangle\right)^2}{N-1}}$	Position fluctuation in Z-direction for force unit *i*
<DRi> ⁑†	$\Delta R_i = \langle \Vert \mathbf{R}_i - \langle \mathbf{R}_i \rangle \Vert \rangle$	Mean value of position fluctuation magnitude on force unit *i*
std(DRi) ⁑†‡	$\sigma_{\Delta R_i} = \sqrt{\frac{\sum_i^N (\Delta R_i - \langle \Delta R_i \rangle)^2}{N-1}}$	Sample standard deviation of position fluctuation magnitude on force unit *i*
<DFiDFi> *†	$\langle \Delta\mathbf{F}_i \Delta\mathbf{F}_i \rangle_{m,n} = \langle \Delta F_{i,m} \Delta F_{i,n} \rangle$ $m,n \in x,y,z$	The average matrix of force fluctuation product on force unit *i*
<DFiDFi>λ *†	$\lambda_k, k \in \alpha,\beta,\gamma$	Eigenvalues of the average matrix of force fluctuation product on force unit *i*
pca(<DFiDFi>) *†	$U_{km}, k \in \alpha,\beta,\gamma$ $m \in x,y,z$	Eigenvectors of the average matrix of force fluctuation product on force unit *i*
<DRiDRi> ⁑†	$\langle \Delta\mathbf{F}_i \Delta\mathbf{F}_i \rangle_{m,n} = \langle \Delta F_{i,m} \Delta F_{i,n} \rangle$ $m,n \in x,y,z$	The average matrix of position fluctuation product on force unit *i*
<DRiDRi>λ ⁑†	$\lambda_k, k \in \alpha,\beta,\gamma$	Eigenvalues of the average matrix of position fluctuation product on force unit *i*
pca(<DRiDRi>) ⁑†	$U_{km}, k \in \alpha,\beta,\gamma$ $m \in x,y,z$	Eigenvectors of the average matrix of position fluctuation product on force unit *i*
<FiDRi> ⁂ †	$\langle \mathbf{F}_i \Delta\mathbf{R}_i \rangle_{m,n} = \langle F_{i,m} \Delta R_{i,n} \rangle$ $m,n \in x,y,z$	The average matrix of force and position fluctuation product on force unit *i*
trace(<FiDRi>) ⁂ †	$\langle \mathbf{F}_i \cdot \Delta\mathbf{R}_i \rangle \equiv trace\langle \mathbf{F}_i \Delta\mathbf{R}_i \rangle$	Mean inner product of the average matrix of force and position fluctuation product on force unit *i*
cov(Fi;Ri) ⁂‡	$\langle \mathbf{F}_i \mathbf{R}_i \rangle - \langle \mathbf{F}_i \rangle \langle \mathbf{R}_i \rangle$	The covariance matrix between the average force and position for force unit *i*
p(Fi;Ri) ⁂‡	$\frac{\langle \mathbf{F}_i \mathbf{R}_i \rangle - \langle \mathbf{F}_i \rangle \langle \mathbf{R}_i \rangle}{\sqrt{\left(\langle \mathbf{F}_i^2 \rangle - \langle \mathbf{F}_i \rangle^2\right)\left(\langle \mathbf{R}_i^2 \rangle - \langle \mathbf{R}_i \rangle^2\right)}}$	The covariance coefficients between the average force and position for force unit *i*
cos(Fi;DRi) ⁂ †	$\left\langle \frac{\mathbf{F}_i \cdot \Delta\mathbf{R}_i}{\sqrt{\Vert\mathbf{F}_i\Vert^2 \Vert\Delta\mathbf{R}_i\Vert^2}} \right\rangle$	Mean cosine value of the angle between force and position fluctuation vectors for force unit *i*
deg(Fi;DRi) ⁂ †	$\left\langle \mathrm{acos} \frac{\mathbf{F}_i \cdot \Delta\mathbf{R}_i}{\sqrt{\Vert\mathbf{F}_i\Vert^2 \Vert\Delta\mathbf{R}_i\Vert^2}} \right\rangle$	Mean value of the angle between force and position fluctuation vectors for force unit *i*

* Force data should be provided with the *atomf* tag specified. Details related to tags used in *readFA* program can be found in Table 7.

⁑ Coordinate data should be provided with the *coord* tag specified.

⁂ Both force data and coordinate data should be provided with the *fcoord* tag specified.

‡ With at least two frames involved in analysis.

† The fluctuation analysis feature should be kept on without —*no-prerun* option specifed.

1.3.2 Pairwise Force Analysis Module

The pairwise force analysis module analyze pairwise forces between force unit pair (i, j) where extremely small pairwise forces are filtered out when analyzing to reduce computational cost. The statistical analysis results of pairwise forces $\mathbf{F}_{ij}$ and displacement vectors $\mathbf{R}_{ij}$ for the force unit pair ij, together with the projection of pairwise force onto the displacement vector, e.g., the vector connecting the two force units, which is denoted as $\boldsymbol{F}_{ij}^{proj}$ are calculated (Table 4).

Table 4 Detailed format of pairwise force analysis results CSV files by *readFA* program on each force unit pairs.

<table>
<tr><th>Data Column</th><th>Expression</th><th>Description</th></tr>
<tr><td>i, j</td><td></td><td>Force unit indexes starting from 0</td></tr>
<tr><td><Fijx> *</td><td>$\langle F_{ij,x} \rangle$</td><td>Mean value of pairwise force X-component on force unit pair ij</td></tr>
<tr><td><Fijy> *</td><td>$\langle F_{ij,y} \rangle$</td><td>Mean value of pairwise force Y-component on force unit pair ij</td></tr>
<tr><td><Fijz> *</td><td>$\langle F_{ij,z} \rangle$</td><td>Mean value of pairwise force Z-component on force unit pair ij</td></tr>
<tr><td><|Fij|> *</td><td>$\langle \|\mathbf{F}_{ij}\| \rangle$</td><td>Mean value of pairwise force scalar on force unit pair ij</td></tr>
<tr><td><|Fij|^2>^1/2 *</td><td>$\sqrt{\langle \|\mathbf{F}_{ij}\|^2 \rangle} = \sqrt{\langle F_{ij,x}^2 + F_{ij,y}^2 + F_{ij,z}^2 \rangle}$</td><td>Square root of mean value of pairwise force vectors norm squared on force unit pair ij</td></tr>
<tr><td><Fij> *</td><td>$\|\langle \mathbf{F}_{ij} \rangle\| = \sqrt{\langle F_{ij,x} \rangle + \langle F_{ij,y} \rangle + \langle F_{ij,z} \rangle}$</td><td>Scalar of mean pairwise force vector on force unit pair ij</td></tr>
<tr><td>std(Fijx) *‡</td><td>$\sigma_{F_{ij,x}} = \sqrt{\frac{\sum_{ij}^{P}\left(F_{ij,x} - \langle F_{ij,x} \rangle\right)^2}{N-1}}$</td><td>Sample standard deviation of pairwise force X-component on force unit pair ij</td></tr>
<tr><td>std(Fijy) *‡</td><td>$\sigma_{F_{ij,y}} = \sqrt{\frac{\sum_{ij}^{P}\left(F_{ij,y} - \langle F_{ij,y} \rangle\right)^2}{N-1}}$</td><td>Sample standard deviation of pairwise force Y-component on force unit pair ij</td></tr>
<tr><td>std(Fijz) *‡</td><td>$\sigma_{F_{ij,z}} = \sqrt{\frac{\sum_{ij}^{P}\left(F_{ij,z} - \langle F_{ij,z} \rangle\right)^2}{N-1}}$</td><td>Sample standard deviation of pairwise force Z-component on force unit pair ij</td></tr>
<tr><td>std(|Fij|) *‡</td><td>$\sigma_{\|\mathbf{F}_{ij}\|} = \sqrt{\frac{\sum_{ij}^{P}\left(\|\mathbf{F}_{ij}\| - \langle \|\mathbf{F}_{ij}\| \rangle\right)^2}{N-1}}$</td><td>Sample standard deviation of pairwise force scalar on force unit pair ij</td></tr>
<tr><td>|std(Fij)| *‡</td><td>$\left\|\sigma_{\mathbf{F}_{ij}}\right\| = \sqrt{\sigma_{F_{ij,x}}^2 + \sigma_{F_{ij,x}}^2 + \sigma_{F_{ij,x}}^2}$</td><td>—</td></tr>
<tr><td><DFij> *†</td><td>$\Delta F_{ij} = \langle \|\mathbf{F}_{ij} - \langle \mathbf{F}_{ij} \rangle\| \rangle$</td><td>Mean value of pairwise force fluctuation magnitude on force unit pair ij</td></tr>
</table>

<table>
<tr><td>std(DFij) *†‡</td><td>$\sigma_{\Delta F_{ij}} = \sqrt{\frac{\sum_{ij}^{P}\left(\Delta F_{ij} - \langle \Delta F_{ij} \rangle\right)^2}{N-1}}$</td><td>Sample standard deviation of pairwise force fluctuation magnitude on force unit pair ij</td></tr>
<tr><td><Rijx> ⁑</td><td>$\langle R_{ij,x} \rangle$</td><td>Mean displacement in X-direction for force unit pair ij</td></tr>
<tr><td><Rijy> ⁑</td><td>$\langle R_{ij,y} \rangle$</td><td>Mean displacement in Y-direction for force unit pair ij</td></tr>
<tr><td><Rijz> ⁑</td><td>$\langle R_{ij,z} \rangle$</td><td>Mean displacement in Z-direction for force unit pair ij</td></tr>
<tr><td><|Rij|> ⁑</td><td>$\langle \lVert \mathbf{R}_{ij} \rVert \rangle \equiv \langle d_{ij} \rangle$</td><td>Mean distance for force unit pair ij</td></tr>
<tr><td><|Rij|^2>^1/2 ⁑</td><td>$\sqrt{\langle \lVert \mathbf{R}_{ij} \rVert^2 \rangle} = \sqrt{\langle R_{ij,x}^2 + R_{ij,y}^2 + R_{ij,z}^2 \rangle}$</td><td>—</td></tr>
<tr><td><Rij> ⁑</td><td>$\lVert \langle \mathbf{R}_{ij} \rangle \rVert = \sqrt{\langle R_{ij,x} \rangle + \langle R_{ij,y} \rangle + \langle R_{ij,z} \rangle}$</td><td>Scalar of mean displacement vector for force unit pair ij</td></tr>
<tr><td>std(Rijx) ⁑‡</td><td>$\sigma_{R_{ij,x}} = \sqrt{\frac{\sum_{ij}^{P}\left(R_{ij,x} - \langle R_{ij,x} \rangle\right)^2}{N-1}}$</td><td>Sample standard deviation of displacement in X-direction for force unit pair ij</td></tr>
<tr><td>std(Rijy) ⁑‡</td><td>$\sigma_{R_{ij,y}} = \sqrt{\frac{\sum_{ij}^{P}\left(R_{ij,y} - \langle R_{ij,y} \rangle\right)^2}{N-1}}$</td><td>Sample standard deviation of displacement in Y-direction for force unit pair ij</td></tr>
<tr><td>std(Rijz) ⁑‡</td><td>$\sigma_{R_{ij,z}} = \sqrt{\frac{\sum_{ij}^{P}\left(R_{ij,z} - \langle R_{ij,z} \rangle\right)^2}{N-1}}$</td><td>Sample standard deviation of displacement in Z-direction for force unit pair ij</td></tr>
<tr><td>std(|Rij|) ⁑‡</td><td>$\sigma_{\lVert \mathbf{R}_{ij} \rVert} = \sqrt{\frac{\sum_{ij}^{P}\left(\lVert \mathbf{R}_{ij} \rVert - \langle \lVert \mathbf{R}_{ij} \rVert \rangle\right)^2}{N-1}}$</td><td>Sample standard deviation of distance for force unit pair ij</td></tr>
<tr><td>|std(Rij)| ⁑‡</td><td>$\lVert \sigma_{\mathbf{R}_{ij}} \rVert = \sqrt{\sigma_{R_{ij,x}}^2 + \sigma_{R_{ij,x}}^2 + \sigma_{R_{ij,x}}^2}$</td><td>—</td></tr>
<tr><td><DRij> ⁑†</td><td>$\Delta R_{ij} = \langle \lVert \mathbf{R}_{ij} - \langle \mathbf{R}_{ij} \rangle \rVert \rangle$</td><td>Mean value of distance fluctuation magnitude for force unit pair ij</td></tr>
<tr><td>std(DRij) ⁑†‡</td><td>$\sigma_{\Delta R_{ij}} = \sqrt{\frac{\sum_{ij}^{P}\left(\Delta R_{ij} - \langle \Delta R_{ij} \rangle\right)^2}{N-1}}$</td><td>Sample standard deviation of distance fluctuation magnitude for force unit pair ij</td></tr>
<tr><td><FijProj> ⁂</td><td>$F_{ij}^{proj} = \frac{\mathbf{F}_{ij} \cdot \mathbf{R}_{ij}}{\lVert \mathbf{R}_{ij} \rVert}$</td><td>Mean value of projection of pairwise force onto displacement vector on force unit pair ij</td></tr>
<tr><td>std(FijProj) ⁂‡</td><td>$\sigma_{F_{ij}^{proj}} = \sqrt{\frac{\sum_{ij}^{P}\left(F_{ij}^{proj} - \langle F_{ij}^{proj} \rangle\right)^2}{N-1}}$</td><td>Sample standard deviation of projection of pairwise force on force unit pair ij</td></tr>
<tr><td><FijRij> ⁂</td><td>$\langle \mathbf{F}_{ij}\mathbf{R}_{ij} \rangle_{m,n} = \langle F_{ij,m} R_{ij,n} \rangle$
$m, n \in x, y, z$</td><td>The average matrix of pairwise force and displacement vector product for force unit pair ij</td></tr>
<tr><td>trace(<FijRij>) ⁂</td><td>$\langle \mathbf{F}_{ij} \cdot \mathbf{R}_{ij} \rangle \equiv trace\langle \mathbf{F}_{ij}\mathbf{R}_{ij} \rangle$</td><td>Mean inner product of the average matrix of pairwise force and displacement vector product for force unit pair ij</td></tr>
<tr><td><DFijDRij> ⁂ †</td><td>$\langle \Delta\mathbf{F}_{ij}\Delta\mathbf{R}_{ij} \rangle_{m,n} = \langle \Delta F_{ij,m} \Delta R_{ij,n} \rangle$
$m, n \in x, y, z$</td><td>The average matrix of pairwise force and displacement fluctuation product for force unit pair ij</td></tr>
<tr><td>trace(<DFijDRij>) ⁂ †</td><td>$\langle \Delta\mathbf{F}_{ij} \cdot \Delta\mathbf{R}_{ij} \rangle \equiv trace\langle \Delta\mathbf{F}_{ij}\Delta\mathbf{R}_{ij} \rangle$</td><td>Mean inner product of the average matrix of pairwise force and displacement fluctuation product for force unit pair ij</td></tr>
</table>

cov(Fij;Rij) ⁂‡	$\langle \mathbf{F}_{ij}\mathbf{R}_{ij} \rangle - \langle \mathbf{F}_{ij} \rangle \langle \mathbf{R}_{ij} \rangle$	The covariance matrix between pairwise force and displacement fluctuation for force unit pair *ij*
p(Fij;Rij) ⁂‡	$\frac{\langle \mathbf{F}_{ij}\mathbf{R}_{ij} \rangle - \langle \mathbf{F}_{ij} \rangle \langle \mathbf{R}_{ij} \rangle}{\sqrt{\left(\langle \mathbf{F}_{ij}^2 \rangle - \langle \mathbf{F}_{ij} \rangle^2\right)\left(\langle \mathbf{R}_{ij}^2 \rangle - \langle \mathbf{R}_{ij} \rangle^2\right)}}$	The covariance coefficients between pairwise force and displacement fluctuation for force unit pair *ij*
cos(Fij;Rij) ⁂	$\left\langle \frac{\mathbf{F}_{ij} \cdot \mathbf{R}_{ij}}{\sqrt{\lVert \mathbf{F}_{ij} \rVert^2 \lVert \mathbf{R}_{ij} \rVert^2}} \right\rangle$	Mean cosine value of the angle between pairwise force and displacement vector for force unit pair *ij*
deg(Fij;Rij) ⁂	$\left\langle \operatorname{acos} \frac{\mathbf{F}_{ij} \cdot \mathbf{R}_{ij}}{\sqrt{\lVert \mathbf{F}_{ij} \rVert^2 \lVert \mathbf{R}_{ij} \rVert^2}} \right\rangle$	Mean value of the angle between pairwise force and displacement vector for force unit pair *ij*
cos(DFij;DRij) ⁂ †	$\left\langle \frac{\Delta\mathbf{F}_{ij} \cdot \Delta\mathbf{R}_{ij}}{\sqrt{\lVert \Delta\mathbf{F}_{ij} \rVert^2 \lVert \Delta\mathbf{R}_{ij} \rVert^2}} \right\rangle$	Mean cosine value of the angle between pairwise force and displacement fluctuation for force unit pair *ij*
deg(DFij;DRij) ⁂ †	$\left\langle \operatorname{acos} \frac{\Delta\mathbf{F}_{ij} \cdot \Delta\mathbf{R}_{ij}}{\sqrt{\lVert \Delta\mathbf{F}_{ij} \rVert^2 \lVert \Delta\mathbf{R}_{ij} \rVert^2}} \right\rangle$	Mean value of the angle between pairwise force and displacement fluctuation for force unit pair *ij*

* Force data should be provided with the *atomf* tag specified.

⁑ Coordinate data should be provided with the *coord* tag specified.

⁂ Both force data and coordinate data should be provided with the *fcoord* tag specified.

‡ With at least two frames involved in analysis.

† The fluctuation analysis feature should be kept on without —*no-prerun* option specifed.

1.3.3 The Correlation Analysis Module

The correlation analysis module computes the correlations between force and force, force and displacement, and displacement and displacement based on results from summed force analysis module. The computed variables include the correlations between forces $\mathbf{F}_i, \mathbf{F}_j$, force fluctuations $\Delta\mathbf{F}_i, \Delta\mathbf{F}_j$, force and position fluctuation, force fluctuation and position fluctuation at force unit *i* and *j*, respectively. The Table 5 lists only results related to the correlations between $\mathbf{F}_i$ and $\mathbf{F}_j$. Other pairs of physical quantities involved in correlation analysis also include $\Delta\mathbf{F}_i, \Delta\mathbf{F}_j$ / $\Delta\mathbf{R}_i, \Delta\mathbf{R}_j$ / $\mathbf{F}_i, \Delta\mathbf{R}_j$ / $\Delta\mathbf{R}_i, \mathbf{F}_j$, and so on.

Table 5 Detailed format of correlation force analysis results CSV files by *readFA* program for every force unit pairs.

Data Column	Expression	Description
i, j		Force unit indexes starting from 0

*<FiFj> **	$\langle \mathbf{F}_i\mathbf{F}_j\rangle_{m,n} = \langle F_{i,m}F_{j,n}\rangle$ $m,n \in x,y,z$	The average matrix of force product for force units *i* and *j*
*trace(<FiFj>) **	$\langle \mathbf{F}_i \cdot \mathbf{F}_j\rangle \equiv trace\langle \mathbf{F}_i\mathbf{F}_j\rangle$	Mean inner product of the average matrix of force product for force units *i* and *j*
*<FiFj>λ **	$\lambda_k, k \in \alpha,\beta,\gamma$	Eigenvalues of the average matrix of force product for force units *i* and *j*
*pca(<FiFj>) **	$U_{km}, k \in \alpha,\beta,\gamma$ $m \in x,y,z$	Eigenvectors of the average matrix of force product for force units *i* and *j*
*cov(Fi;Fj) *‡*	$\langle \mathbf{F}_i\mathbf{F}_j\rangle - \langle \mathbf{F}_i\rangle\langle \mathbf{F}_j\rangle$	The covariance matrix between average forces for force units *i* and *j*
*p(Fi;Ri) *‡*	$\frac{\langle \mathbf{F}_i\mathbf{F}_j\rangle - \langle \mathbf{F}_i\rangle\langle \mathbf{F}_j\rangle}{\sqrt{\left(\langle \mathbf{F}_i^2\rangle - \langle \mathbf{F}_i\rangle^2\right)\left(\langle \mathbf{F}_j^2\rangle - \langle \mathbf{F}_j\rangle^2\right)}}$	The covariance coefficients between average forces for force units *i* and *j*
*cos(Fi;Fj) **	$\left\langle \frac{\mathbf{F}_i \cdot \mathbf{F}_j}{\sqrt{\lVert \mathbf{F}_i \rVert^2 \lVert \mathbf{F}_j \rVert^2}} \right\rangle$	Mean cosine value of the angle between forces for force units *i* and *j*
*deg(Fi;Fj) **	$\left\langle \operatorname{acos} \frac{\mathbf{F}_i \cdot \mathbf{F}_j}{\sqrt{\lVert \mathbf{F}_i \rVert^2 \lVert \mathbf{F}_j \rVert^2}} \right\rangle$	Mean value of the angle between forces for force units *i* and *j*

* Force data should be provided with the *atomf* tag specified.

‡ With at least two frames involved in analysis.

1.3.4 The Network Analysis Module

readFA program supports analysis of pairwise forces between residues in protein. The pairwise forces as bridges connect residues as a whole network. The properties of protein internal residue force network are useful in studying properties of protein response mechanisms. The mean force scalar $\lVert \mathbf{F}_{ij} \rVert$ is used to determine edges and their weights in an undirected graph:

$$e_{ij} = \begin{cases} 0, & \lVert \mathbf{F}_{ij} \rVert < F_{cutoff}^{graph} \\ 1, & \lVert \mathbf{F}_{ij} \rVert \geq F_{cutoff}^{graph} \end{cases}, \quad |E| = \sum_{i,j} e_{ij}$$

$$w_{ij} = \frac{\left(\prod_{e_{ij}=1} \lVert \mathbf{F}_{ij} \rVert\right)^{\frac{1}{|E|}}}{\lVert \mathbf{F}_{ij} \rVert} \quad (1)$$

1.3.5 Regression Analysis

Regression is used to explore the relationship of pairwise force projection changes versus the distance in between. Usually, the mean pairwise force vectors are not in a line of the displacement vector connecting the residues pair. Due to the deviation between the force and displacement vectors, the projection of $\mathbf{F}_{ij}$ onto the corresponding displacement vector, F_{ij}^{proj}, is chosen for analysis. In addition, F_{ij}^{proj} is a scalar with sign. The positive value of F_{ij}^{proj} shows the attractive tendency of two residues and the negative sign indicates the repulsive tendency. It is worthy noting that the tendency may not result into motions, since the tendency of movement of one residue pair is also limited by restraints from other residue pairs. Due to unknown measurement errors in F_{ij}^{proj} and d_{ij}, and the large difference in variances, the Deming regression method was adopted. The variable d_{ij} is chosen to be the predictor variable, while the F_{ij}^{proj} as the response variable. The regression equation can be expressed as follows:

$$\begin{cases} Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t \\ X_t = x_t + \mu_t \end{cases} \tag{2}$$

where X_t, Y_t are the measurement value of predictor and response variables, respectively. x_t is the true value of predictor variable. ε_t, μ_t correspond to the errors of the predictor and response variables, respectively, and β_0, β_1 are the intercept and slope in the regression expression, respectively.

The sample covariance matrix is as follows:

$$M = \begin{bmatrix} \langle X_t^2 \rangle - \langle X_t \rangle^2 & \langle X_t Y_t \rangle - \langle X_t \rangle \langle Y_t \rangle \\ \langle X_t Y_t \rangle - \langle X_t \rangle \langle Y_t \rangle & \langle Y_t^2 \rangle - \langle Y_t \rangle^2 \end{bmatrix} \equiv \begin{bmatrix} M_{XX} & M_{XY} \\ M_{XY} & M_{YY} \end{bmatrix} \tag{3}$$

In common cases where $M_{XY} \neq 0$, the estimated slope and intercept are:

$$\hat{\beta}_1 = \frac{M_{YY} - \delta M_{XX} + [(M_{YY} - \delta M_{XX})^2 + 4\delta M_{XY}^2]^{\frac{1}{2}}}{2M_{XY}}$$

$$\hat{\beta}_0 = \langle Y_t \rangle - \hat{\beta}_1 \langle X_t \rangle \tag{4}$$

where δ is the ratio of variances of the response variable to the predictor variable, $\delta = s_{yy}/s_{xx}$.

The estimated values for the predictor variable and its error variance are shown as follows:

$$\hat{\sigma}_{XX}^2 = -\frac{M_{YY} - \delta M_{XX} - [(M_{YY} - \delta M_{XX})^2 + 4\delta M_{XY}^2]^{\frac{1}{2}}}{2\delta}$$

$$\hat{\sigma}_u^2 = \frac{M_{YY} + \delta M_{XX} - [(M_{YY} - \delta M_{XX})^2 + 4\delta M_{XY}^2]^{\frac{1}{2}}}{2M_{XY}}$$

（5）

Based on the estimated slope and the approximate distribution of the error, the estimated value of the covariance matrix is given by:

$$\hat{V}\{\hat{\beta}_0, \hat{\beta}_1\} = \begin{bmatrix} \frac{S_{vv}}{n} + \langle X_t \rangle^2 \hat{V}\{\hat{\beta}_1\} & -\langle X_t \rangle \hat{V}\{\hat{\beta}_1\} \\ -\langle X_t \rangle \hat{V}\{\hat{\beta}_1\} & \hat{V}\{\hat{\beta}_1\} \end{bmatrix}$$

（6）

where

$$\hat{V}\{\hat{\beta}_1\} = \frac{1}{n-1}\left[\frac{S_{vv}}{\hat{\sigma}_{XX}} + \left(\frac{\hat{\sigma}_u}{\hat{\sigma}_{XX}}\right)^2 S_{vv} - \left(\frac{\hat{\beta}_1 \hat{\sigma}_u^2}{\hat{\sigma}_{XX}}\right)^2\right]$$

$$S_{vv} = \frac{n-1}{n-2}(\delta + \hat{\beta}_1^2)\hat{\sigma}_u^2$$

（7）

And the confidence interval for the slope can be estimated as:

$$\hat{\beta}_1 \pm Z_{\frac{\alpha}{2}}\left[\hat{V}\{\hat{\beta}_1\}\right]^{\frac{1}{2}}$$

（8）

As well as the confidence interval for the intercept:

$$\hat{\beta}_0 \pm Z_{\frac{\alpha}{2}}\left[\frac{S_{vv}}{n} + \langle X_t \rangle^2 \hat{V}\{\hat{\beta}_1\}\right]^{\frac{1}{2}}$$

（9）

The results of the Deming regression analysis implemented in *readFA* program are listed in the table (Table 6). These results are integrated as columns in pairwise force analysis results.

Table 6 Estimation results by Deming regression method by *readFA* program.

Data Column	**Expression**	**Description**
i, j		Force unit indexes starting from 0
*var(FijProj)/var(|Rij|) **	$\delta = \frac{\sigma^2\left(F_{ij}^{proj}\right)}{\sigma^2\left(\left\|R_{ij}\right\|\right)}$	Ratio of variances of F_{ij}^{proj} to d_{ij}

b *	$\hat{\beta}_0$	Estimated intercept of Deming regression
k *	$\hat{\beta}_1$	Estimated slope of Deming regression
var(E(|Rij|)) *	$\hat{\sigma}^2(\lvert R_{ij}\rvert)$	Estimated variance of d_{ij}
var(E(FijProj)) *	$\hat{\sigma}^2(F_{ij}^{proj})$	Estimated variance of F_{ij}^{proj}
var(Eerr(|Rij|)) *	$\hat{\sigma}_{err}^2(F_{ij}^{proj})$	Estimated error variance of d_{ij}
var(Eerr(FijProj)) *	$\hat{\sigma}_{err}^2(\lvert R_{ij}\rvert)$	Estimated error variance of F_{ij}^{proj}
lb(CI95(b)) *	$\hat{\beta}_0 - Z_{0.95}$	The lower bound of the 95% confidence interval for the intercept
ub(CI95(b)) *	$\hat{\beta}_0 + Z_{0.95}$	The upper bound of the 95% confidence interval for the intercept
lb(CI95(k)) *	$\hat{\beta}_1 - Z_{0.95}$	The lower bound of the 95% confidence interval for the slope
ub(CI95(k)) *	$\hat{\beta}_1 + Z_{0.95}$	The upper bound of the 95% confidence interval for the slope

* Force and coordinate data should be provided with both *coord* and *fcoord* tags specified. Additionally, there should be at least two frames involved in analysis.

2. How to Use READFA Program

2.1 Command Line Options

The input, data processing methods, and output results of *readFA* program are controlled by command line options. All command line options start with a hyphen (-) or a dash (–), following one or more arguments. Option keywords and arguments are separated by space(s).

2.1.1 Command Line Options for Controlling Analysis Modes

The *--out* option sets the analysis mode of *readFA* program. This option has the lowest priority, so the detailed analysis mode would be modified by other related command line options and input data types. The analysis mode can be described as a combination of different statistical analysis operations, and these operations can be described with a set of tags. Analysis tags include **raw** (no statistics, only extract raw data), **atomf** (related to summed force analysis), **pf** (related to pairwise force analysis), **coord** (related to coordinate analysis), **fcoord** (related to force and coordinate correlation analysis), and **fgraph** (related to network analysis). The detailed relationship between tags and detailed analysis modes is shown in Table 7.

Table 7 Details of tags used by each analysis mode specified by keywords.

Analysis Mode Keyword	**Tags (Included tags in analysis mode are shown and checked)**					
raw	*raw*✓					
summed		*atomf*✓				
pairwise			*pf*✓			
coord				*coord*✓		
forcecoord					*fcoord*✓	
force-graph						*fgraph*✓
force		*atomf*✓	*pf*✓			
no-pairwise		*atomf*✓		*coord*✓	*fcoord*✓	
no-forcecoord		*atomf*✓	*pf*✓	*coord*✓		
all		*atomf*✓	*pf*✓	*coord*✓	*fcoord*✓	*fgraph*✓

2.1.2 Command Line Options for Controlling Input

Table 8 Command-line options and arguments for controlling input.

Command Line Options	**Optional or Required**	**Number of Arguments**	**Description**	**Comments**
-f	required	1	Path to the force data file.	Force data files are usually binary files with the suffixes .far or .for, generated by *GROMACS-FA*.
--map	required	2	Paths of index map files for GROUP A and GROUP B in analysis group pairs. The argument corresponding to GROUP A must be as first.	The force unit index recorded in force data file is reindexed and remapped via the given index map files provided here.
--ref	optional	1	Path to the coordinate file of the reference system.	If coordinate data for the reference system is not provided, the rotation correction feature would be disabled automatically.
--coord	optional	2	Paths of coordinate files for GROUP A and GROUP B in analysis group pairs. The argument corresponding to GROUP A must be as first.	If coordinate data is not provided, the analysis features corresponding to the *coord/fcoord/fgraph* tags will be disabled automatically.

2.1.3 Command Line Options for Controlling Output

Table 9 Command-line options and arguments for controlling output results.

Command Line Options	Optional or Required	Number of Arguments	Description	Comments
-o	required	1	Output result filename.	The output result file will be in Comma-Separated Values (.csv) format, and the filename should include the suffix. There might be issues with filenames containing directories in prefix. It is recommended to run *readFA* program in the working directory
--out-digits	optional	1	Specify number of decimal places in results	The output result format is in decimal form. By default, the analysis results are rounded to 3 decimal places.

2.1.4 Command Line Options for Controlling Data Processing Methods

Table 2 Command-line options and arguments for controlling data processing methods.

Command Line Options	Optional or Required	Number of Arguments	Description	Comments
--out	required	1	Specify the keyword for controlling the analysis mode, see Table 7 for details.	See section 2.1.1 Command Line Options for Controlling Analysis Modes.
--block	optional	2	Specify the start and end frame indexes of interesting segments	The start and end frame specified by indexes are included in range. If the end index is -1, it is automatically converted to the end frame of trajectory. The start index should be no less than 0. If the start index is greater than the end index, or the total frames contained in the trajectory, the analysis segment will be empty. If this option is not specified, it is default to analyze the entire trajectory input
--force-graph-edge	optional	1	Specify the force threshold for identifying a given force component pair as an edge in the graph	Typically, whether i - j constitutes an edge in the graph is determined by the absolute value of the mean of $\left\|\mathbf{F}_{ij}\right\|$, with the unit being pN. If this option is not specified, the default threshold is set to 0.1 pN

--no-forcerot	optional	0	—	With this option specified, the rotation correction feature would be disabled automatically
--no-prerun	optional	0	—	With this option specified, the mean values of forces and coordinates would not be pre-calculated, and the fluctuation analysis features that depend on mean values would be disabled automatically
--no-crossanal	optional	0	—	With this option specified, the correlation analysis feature would be disabled automatically. It would be a large memory and storage cost for correlation analysis with lots of force units, and it is recommended to turn off this feature in this case

2.2 Usage Example

To do a force analysis on simulation system named *pro_md* with force data files available. Assumed that there are force files named *pro_md.for*, index map files for two groups in one analysis group pair *pro_md_grpA.map*, *pro_md_grpB.map* and coordinate files for the reference system and two groups *pro_md_ref.xyz, pro_md_grpA.xyz, pro_md_grpB.xyz* in the working directory. Both force data and coordinate data files have 1001 blocks/frames, which are indexed starting from 0. Therefore the frame indexes are ranged from 0 to 1000.

To analyze pairwise force data file *pro_md.for* for the entire trajectory.

Shell
```
readFA -f pro_md.for -o pro_md_pf.csv --map pro_md_grpA.map pro_md_grpB.map
--ref pro_md_ref.xyz --coord pro_md_grpA.xyz pro_md_grpB.xyz --out all
```

To analyze the data for the first 100 block:

Shell
```
readFA -f pro_md.for -o pro_md_pf.csv --map pro_md_grpA.map pro_md_grpB.map
--ref pro_md_ref.xyz --coord pro_md_grpA.xyz pro_md_grpB.xyz --out all –
block 0 99
```

To extract raw force data from the given data file *pro_md.for*:

Shell
```
readFA -f pro_md.for -o pro_md_raw.csv --map pro_md_grpA.map
```

```
pro_md_grpB.map --out raw
```