Supplementary Methods

3.3. Evaluating pathway projections

The eBDIMS server tries to select by default the best two PCs or NMs for projection i.e. the first two PCs, which typically capture >70% of the structural variance, or the two NMs better overlapped with the modelled transition, which ideally should capture >80% of the movements involved (*Table S1* and (Orellana *et al.*, 2016, 2010)).

Whether these axes adequately capture the motions of interest to understand a transition, needs to be ultimately judged by the user taking into account biological information about the structures. As a rule of thumb, when several distinct experimental conformations are available, PCs are better descriptors than NMs, which typically only track accurately simple rigid-body harmonic motions (illustrated in *Fig.S1*). Note also that NMs computed from different reference structures can be remarkably different (e.g. those computed from open conformations usually track better large-scale transitions). Thus, when analysing transition projections, the user needs to consider two important features:

- **Whether the 2D-projection is effectively separating distinct known functional states** onto different clusters, e.g. structures bound/unbound, solved in similar conditions, etc. If a pair of motion axes group together structures solved in similar conditions (bound to certain ions or ligands, solved at the same pH, etc) or representing the same state (open/closed, inactive/active, bound/unbound, etc) indicates they are relevant to analyze the transition.

- **Whether the 2D-projection of the eBDIMS path is smooth**: since each eBDIMS frame necessarily resembles the preceding and the following one, and all of them progress towards a target structure, in a significant motion axis they will project following a smooth line. Backtracking, sharp changes in direction or zigzag paths indicate the axis does not capture adequately the features of the structures, and cab be caused by either low overlap with the transition (for NMs) or insufficient or irrelevant sampling in the ensemble (for PCs) e.g. when input conformations are not different enough or when N- or C- termini fluctuations dominate the largest variance PCs.

The features of the projected structures should change gradually along significant modes e.g. rMSD versus the target or start structure, collective variables such as pore radius, key interresidue distances, etc. In such ideal 2D-projections, intermediate states will appear on top of or near the eBDIMS paths, which can even seem to depart from a smooth line towards their direction; in cases of large ensembles, intermediates can even appear as series of aligned points following one of the paths (e.g. GLIC locally closed structures, (Orellana *et al.*, 2016)). If a structure cluster is very distant from the low-energy eBDIMS area, it will likely correspond to structures in an entirely different set of biochemical conditions, not expected to be found by eBDIMS or by unbiased MD. Such lateral (not on-pathway) clusters can require high-energy co-substrates like ATP, special ions (Mg^{2+}), or pH/temperature conditions that favour non-spontaneous conformational changes.

In general, the default modes presented by the server are the most representative. However, in the case of complex transitions where small-scale higher modes are involved, or whenever the presented axes do not provide smooth projections and neat clustering, exploration of other modes can be useful:

- For PCs, modes beyond the 2rd can still capture significant variance, and thus be relevant to further understand/classify the motions driving the transition; in general, a PC with variance > 10% should be explored

- For NMs, lower-energy modes can provide better projections than other modes of higher frequency, which can display marginally better overlap with the transition; in cases of two modes with similar overlap, since their motions are orthogonal, it can happen than one is relevant for the transition and the other is not (e.g. shows backtracking, etc)

If no relevant modes are found, we recommend using as reference for NM calculation another structure (typically, NMs from “open” structures render better overlaps with conformational transitions than their “closed” counterparts), and in the case of PC calculations, removing parts of the structure that can introduce noise (like flapping termini) and or changing ensemble composition, which can be distorted e.g. by structures with artifactual changes not related to the transition of interest.

Examples illustrating the above points are provided in *Fig.S1* and in the online documentation.

Supplementary Figures

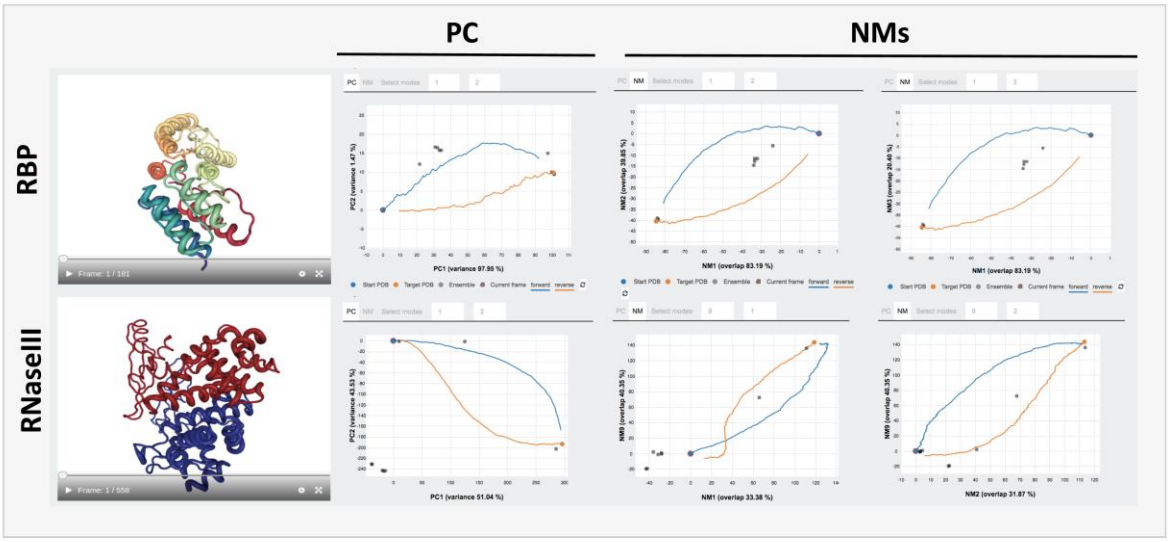

Figure S1. Representative eBDIMS results comparing PCs versus NMs projections. Screenshots of eBDIMS results for RBP (*top*) and RNaseIII (*bottom*) to illustrate the performance of 2D-projections in NM or PCs depending on transition complexity. For simple harmonic transitions like RBP hinge opening (top row), NM-space and PC-space representations both project the transition intermediates as two clusters along the first motion axis (blue); note that in fact, one single mode is enough to cluster the structures since both for PC1 and NM1 greatly concentrate the description of the conformational change (with total variance 97% and overlap 83%, respectively) and thus the second axes has little influence in partition of the structures e.g. NM2 or NM3 render similarly good projections. On the contrary, for the complex sequential motion explored by RNaseIII (bottom row), only PCs can distinguish the four known structural clusters along the catalytic cycle, while NMs, which have low overlap with the sequential movements required, only can separate three of them. According with the complex transition, the best-overlapped mode has high frequency (NM9=40%); note that out of the two next modes better overlapped (~30%), the one rendering the smoothest projection has again higher frequency than the lowest-frequency one, which renders a zig-zag unrealistic eBDIMS projection.

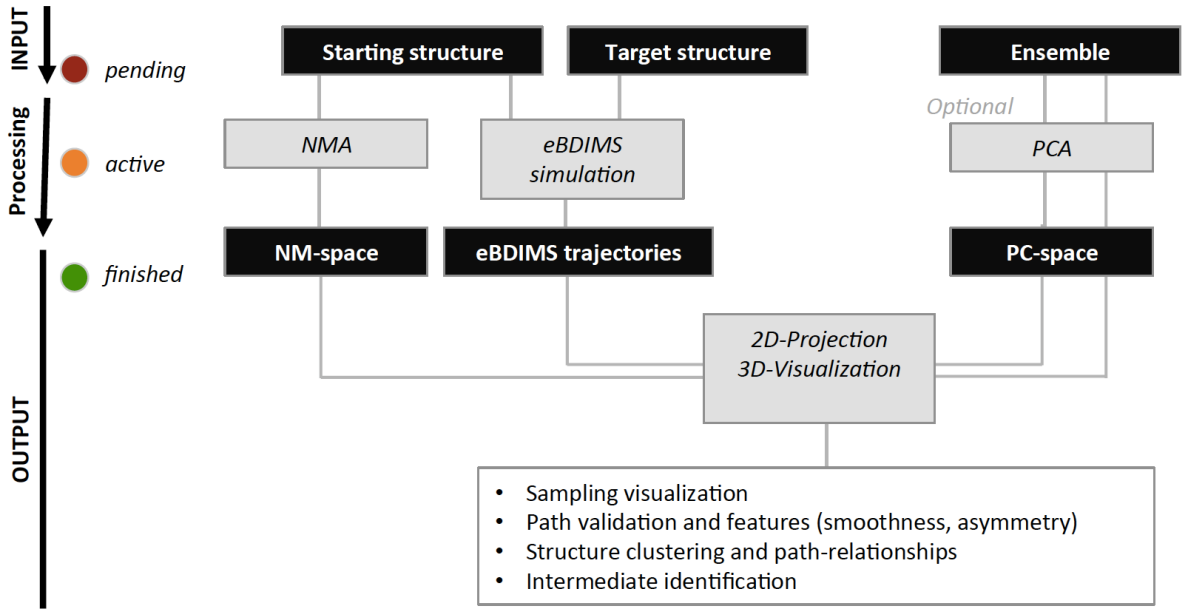

Figure S2. Scheme of the eBDIMS server workflow. Flowchart illustrating the input requirements (end-states and ensemble pdb files), the calculations performed and the output obtained (eBDIMS trajectories and projections onto 2D-motion space).

Supplementary Tables

Table S1. Summary of eBDIMS benchmark results available online. *Nres= Number of residues, n-mer=number of oligomers, Nens=Number of ensemble structures. Note that all examples can achieve lower rMSD values running locally.*

Name	Start	Target	Nres (n-mer)	Initial rMSD	Final rMSD	eBDIMS Server Time	NMA* □□□□□□□□□□□□□□□□□□	Nens	Variance PC1-PC2
RBP$	**1ba2**	**2dri**	270 (1)	6.2	0.80 0.83	120s 101s	0.83(1), 0.4(2)	11	97%-2%
5NTase$	**1oid**	**1hpu**	524 (1)	9.3	0.99 0.98	564s 605s	0.5(3), 0.3(8)	16	94%-4%
RNaseIII$	**1yyo**	**1yyw**	436(2)	17.1	2.1 1.6	634s 568s	0.4(9), 0.3(1)	11	51%-43%
SERCA$	**2c9m**	**1t5s**	994(1)	14.0	2.06 1.86	4291s 4927s	0.53(2), 0.52(3)	65	57%-28%
GLIC$,	**4npq**	**4hfi**	1550(5)	2.6	0.43 0.46	14400s 14400s	0.71(4), 0.24(3)	46	42%-29%
Calmodulin	**1cll**	**3ewv**	134(1)	14.8	1.9 3.7	46s 57s	0.51(4), 0.4(2)	44	41%-27%
mRNA capping enzyme	**1ckmB**	**1ckmA**	330(1)	3.5	0.52 0.43	92s 85s	0.51(5), 0.32(2)	-	-
Human Integrin	**6avu**	**6avq**	1299(2)	31.9	9.1 15.1	14400s 14400s	0.2(7), 0.2(10)	3	94%-6%
Adenylosuccinate synthetase	**1qf5**	**1hoo**	431(1)	2.0	0.26 0.21	333s 1062s	0.51(3), 0.16(6)	-	-
Importin	**2q5dB**	**2q5dA**	859(1)	4.6	0.45 0.41	1660s 1442s	0.54(1), 0.28(7)	8	83%-8%
GROEL	**1oelA**	**1sx4A**	547(1)	12.2	2.2 1.6	728s 775s	0.65(3), 0.40(1)	-	-
Adenylate Kinase	**1ake**	**4ake**	214(1)	18.4	2.4 2.2	620s 607s	0.55(6), 0.46(5)	-	-
Guanylate Kinase	**1ex6**	**1ex7**	186(1)	3.6	0.53 0.42	26s 32s	0.89(1), 0.2(2)	-	-
Maltodextrin Binding Protein	**1omp**	**1anf**	370(1)	3.8	0.42 0.52	201s 159s	0.81(1), 0.47(2)	-	-
Human Lactoferrin	**1lfg**	**1lfh**	691(1)	6.4	0.74 0.77	945s 1030s	0.50 (5), 0.3(3)	-	-
LAO-Binding Protein	**2lao**	**1lst**	238(1)	4.7	0.55 0.56	59s 59s	0.93(1), 0.2(2)	-	-
Aspartate Carbamoyl Transferase	**5at1**	**8atc**	310(2) 153(2)	4.9	0.66 0.62	1855s 1786s	0.55(4), 0.3(3)	-	-
Oligopeptide Binding Protein	**1rkm**	**2rkm**	517(1)	3.1	0.30 0.37	706s 634s	0.96(1), 0.10(3)	-	-

** Defined as* $\alpha_k = \frac{\Delta r_{t-0} \cdot M_k}{\|\Delta r_{t-0}\| \|M_k\|}$ *(0=start, t=target)*

$ Example discussed in Orellana et al. (2016)