Discrete Molecular Dynamics Distinguishes Native-like Binding Poses from Decoys in Difficult Targets

Elizabeth A. Proctor[*,†], Shuangye Yin[‡], Alexander Tropsha[*,†,§,¶], and Nikolay V. Dokholyan[*,†,‡,¶,‖]

[*]Curriculum in Bioinformatics and Computational Biology; [†]Program in Molecular and Cellular Biophysics; [‡]Department of Biochemistry and Biophysics, School of Medicine; [§]Division of Medicinal Chemistry and Natural Products, UNC Eshelman School of Pharmacy; [¶]Center for Systems and Computational Biology; University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States of America

[‖]To whom correspondence should be addressed. Email: dokh@unc.edu

Supplementary Material

We find that all cases in which our method does not successfully distinguish the native pose from the decoys are from the kinase family of proteins. Because kinases are known to be highly flexible and to undergo conformational changes upon the binding of ligands, we attempt to define a metric based on the concept of protein flexibility and correlated motion that will determine whether our method will be successful for a given target. Allosteric change seems an especially likely explanation because of the harmonic constraints that we place on the target backbone during simulation. In order to study backbone movement and correlated motion within the proteins, we conduct unconstrained simulations of two success cases and three non-success cases. We perform 500,000 time steps (~2.5 ns) of simulation for each target. We examine several parameters, described below, however, we are not successful in finding a parameter by which we can separate success from non-success cases.

Average energy of ligand-target complex
Average complex potential energy is recorded every time step (~50 fs). We average these energies over the entire simulation in order to obtain the average energy for the target. We observe no trend in average energy that may distinguish success cases from non-success cases.

Target	Success/Non-success	Average energy (kcal/mol)
Pantothenate synthetase	Success	-1184
JNK3	Success	-1504
P38 kinase	Success	-1471
CHK1	Non-success	-1138
Pim-1 kinase	Non-success	-1210

Root mean square deviation (RMSD) of protein backbone
The RMSD of the protein backbone is measured using the alpha-carbon coordinates from each simulation snapshot, which are collected every 100 time steps (~5 ns). We calculate the backbone RMSD both as a trajectory, using the initial conformation as a reference and as a pairwise distribution of RMSDs (Figure S3). In neither case do we see successes and non-successes partitioned to larger or smaller values, nor a pattern in the shape of the trajectories or distributions, that might explain differences in dynamics between success and failure cases.

Root mean square deviation (RMSD) of pocket residues
The RMSD of pocket residues is measured using the alpha-carbon coordinates for each pocket residue. Pocket residues are determined using the SCREEN2 web-server (1). We calculate the pairwise RMSD distribution of all snapshots to all snapshots. Again, we observe no common trend in shape or values in non-success cases that distinguishes them from success cases (Figure S4).

Binding pocket surface area

We submit each target to the SCREEN2 server (1) (http://luna.bioc.columbia.edu/honiglab/screen2/cgi-bin/screen2.cgi). SCREEN2 locates all cavities on the surface of the protein, and reports various physical and chemical properties, including the residues that form the cavity and the surface area. We verify that the largest identified cavity is the binding pocket of the ligand, and extract the surface area information.

Target	Success/Non-success	Binding pocket surface area (Å)
Acetylcholine esterase	Success	554.9
Pantothenate synthetase	Success	1352.1
JNK3	Success	1395.5
LmrR	Success	110.5
Tuberculosis thymiylate kinase	Success	No significant pockets found
P38 kinase	Success	903.4
CHK1	Non-success	465.9
Pim-1 kinase	Non-success	663.2

In one case, tuberculosis thymiylate kinase (PDB ID: 1W2G), SCREEN2 does not locate any cavity. Visual inspection of the protein surface does not reveal an immediate rationale for this result, since the binding pocket of tuberculosis thymiylate kinase is visually very similar to the binding pockets of the other targets. In the remaining targets, the binding pocket surface areas tend to vary widely, and there is no delineation between success and non-success binding pockets.

Ligand-target contacts

At each movie frame, we calculate the number of ligand-target contacts in two ways. We count any heavy atom of the target that comes within 4.5 Å of any heavy atom of the ligand as a contact. We consider both how many target residues are in contact with the ligand and how many target atoms are in contact with the ligand in each snapshot (Figure S5). The trends in contacts do not differ between residue contacts and atom contacts, so we may assume that there is no large fluctuation in the amount of contacts per residue.

Correlated motion calculations

We calculate the set of pairwise correlation coefficients, C_{ij}, between the alpha-carbons of all residues *i* and *j* of each target as in Sharma et al. (2). Correlation coefficient values range from -1 to 1, with more negative values indicating increasingly anti-correlated motion, and positive values indicating increasingly correlated motion. Patterns vary across success and non-success cases, and no immediate distinguishing factor is apparent (Figure S6).

Correlation metrics

We construct metrics to discern differences between success and non-success cases in the value or magnitude of correlated motion of pocket residues with the rest of the protein. Metrics are a sum of all correlation coefficients C_{ij}, where residue *i* is a pocket residue. We investigate four types of metrics. In the first, we simply add the values of the correlation coefficients. In the second, we add the absolute values of the correlation coefficients, in order to take into account only the magnitudes of correlation found, and not their directions. In the third, we look at only positive correlations, and in the fourth, only negative correlations. In all types, we use two different types of normalization: by the number of pocket residues (*n*), and by the number of correlations in the sum (*N*).

Sums	Normalization by n		Normalization by N	
	Success	*Non-success*	*Success*	*Non-success*
C_{ij} values	2.347		0.000818	
	3.652	13.559	0.0106	0.0498
	1.055	1.72	0.00312	0.00632
C_{ij} magnitudes	46.082		0.16	
	76.87	49.363	0.223	0.181

	81.717	48.792	0.241	0.179
Positive C_{ij}	24.215		0.178	
	40.261	31.461	0.239	0.212
	41.386	25.256	0.255	0.200
Negative C_{ij}	-21.867		-0.144	
	-36.609	-17.901	-0.208	-0.144
	-40.331	-23.535	-0.229	-0.16

Calculation of average minimal path

We define a graph, using all residues as vertices. We impose various cutoffs in the magnitude of the correlation coefficient, ranging from zero to 0.9, for the creation of an edge between any two vertices; a cutoff of zero would result in a complete graph. We calculate the minimal path, L_{ij}, between pocket residue *i* and non-pocket residue *j*. We calculate the average minimal path to non-pocket residues, L_i, for each pocket residue *i*, and the average minimal path for the target as the average of all L_i, $\langle\langle L \rangle\rangle$:

$$\langle L_i \rangle = \frac{1}{N} \sum_{\{j\}} L_{ij}$$

$$\langle\langle L \rangle\rangle = \frac{1}{M} \sum_{\{i\}} \langle L_i \rangle$$

Where *{j}* is the set of non-pocket residues, *{i}* is the set of pocket residues, *N* is the total number of non-pocket residues, and *M* is the total number of pocket residues. We perform these calculations both for an unweighted graph and for a weighted graph, where the correlation coefficients C_{ij} are used as the set of edge weights. We find that the average minimum path from pocket residues to other regions of the protein cannot distinguish between success and non-success cases. We also find no relation between success and the edge cutoff at which the graph becomes disconnected, resulting in no path.

Unweighted								
System		**Cutoffs**						
Target	**Success/Non-success**	*0*	*0.1*	*0.2*	*0.3*	*0.4*	*0.5*	*0.6*
Pantothenate synthetase	Success	1	1.417	1.728	2.157	2.957	4.461	----
JNK3	Success	1	1.295	1.554	1.793	2.255	----	----
P38 kinase	Success	1	1.387	1.678	2.130	3.109	----	----
CHK1	Non-success	1	1.281	1.522	1.756	2.190	2.911	----
Pim-1 kinase	Non-success	1	1.339	1.633	2.023	2.917	----	----
Weighted								
System		**Cutoffs**						
Target	**Success/Non-success**	*0*	*0.1*	*0.2*	*0.3*	*0.4*	*0.5*	*0.6*
Pantothenate synthetase	Success	0.845	0.904	1.001	1.138	1.335	1.679	----
JNK3	Success	0.785	0.807	0.848	0.903	0.993	----	----
P38 kinase	Success	0.831	0.891	0.996	1.140	1.401	----	----
CHK1	Non-success	0.765	0.786	0.824	0.875	0.952	1.070	----
Pim-1 kinase	Non-success	0.816	0.857	0.932	1.054	1.315	----	----

References

1. Nayal M & Honig B (2006) On the nature of cavities on protein surfaces: application to the identification of drug-binding sites. *Proteins* 63(4):892-906.
2. Sharma S, Ding F, & Dokholyan NV (2007) Multiscale modeling of nucleosome dynamics. *Biophys J* 92(5):1457-1470.

Figure S1. AutoDock does not succeed in identifying native pose in difficult targets. AutoDock score for each of the 1000 MedusaDock-generated poses and native pose in: (A) pantothenate synthetase, (B) AChE, (C) JNK3, (D) LmrR, (E) tuberculosis thymidylate kinase, (F) MAP kinase 14, (G) pim-1 kinase. The black open circle and dashed line indicate the native pose and its AutoDock score. CHK1 was unable to be processed by AutoDock as listed in the PDB.

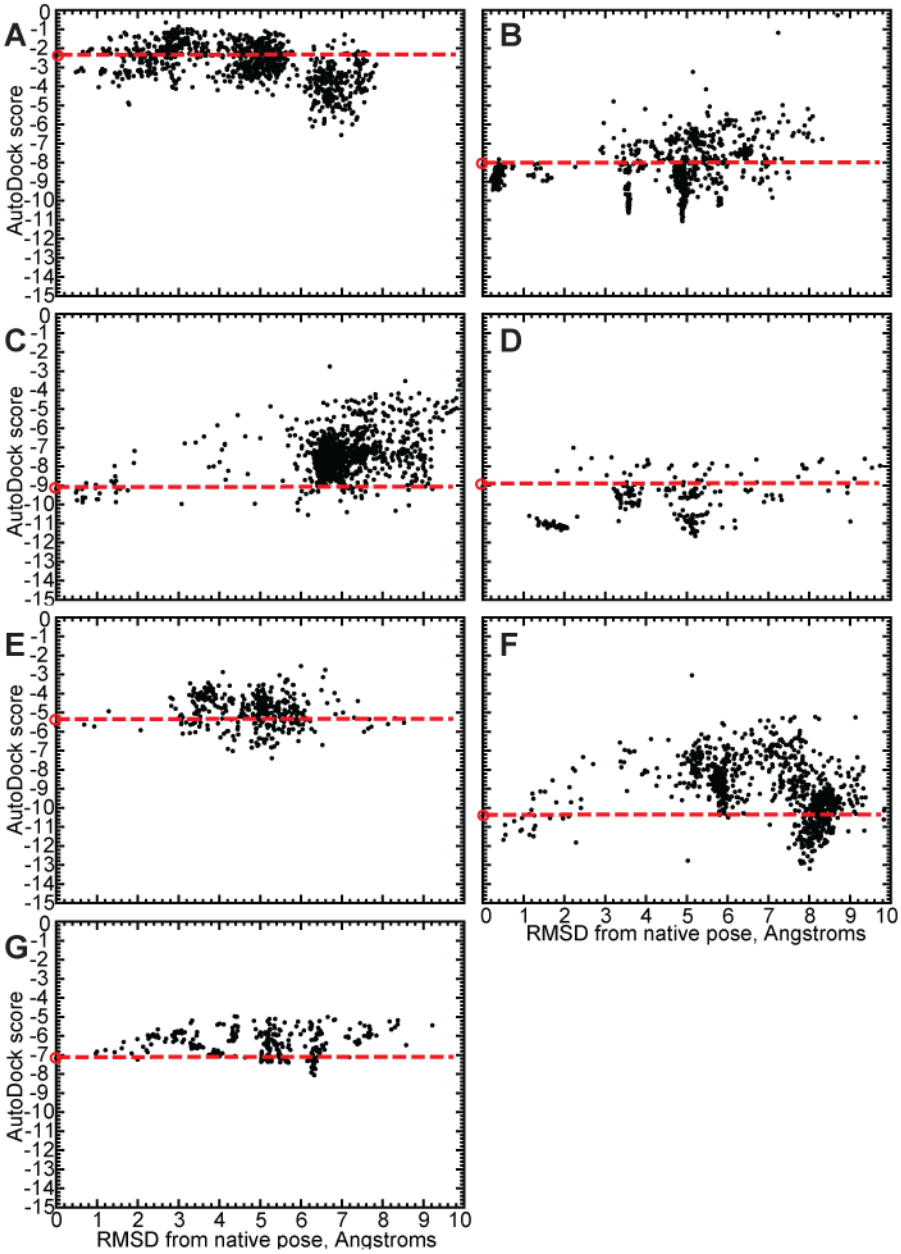

Figure S2. Glide does not succeed in identifying native pose in difficult targets. GlideScore for each of the 1000 MedusaDock-generated poses and native pose in: (A) pantothenate synthetase, (B) AChE, (C) JNK3, (D) LmrR, (E) tuberculosis thymidylate kinase, (F) MAP kinase 14, (G) CHK1, (H) pim-1 kinase. The black open circle and dashed line indicate the native pose and its GlideScore.

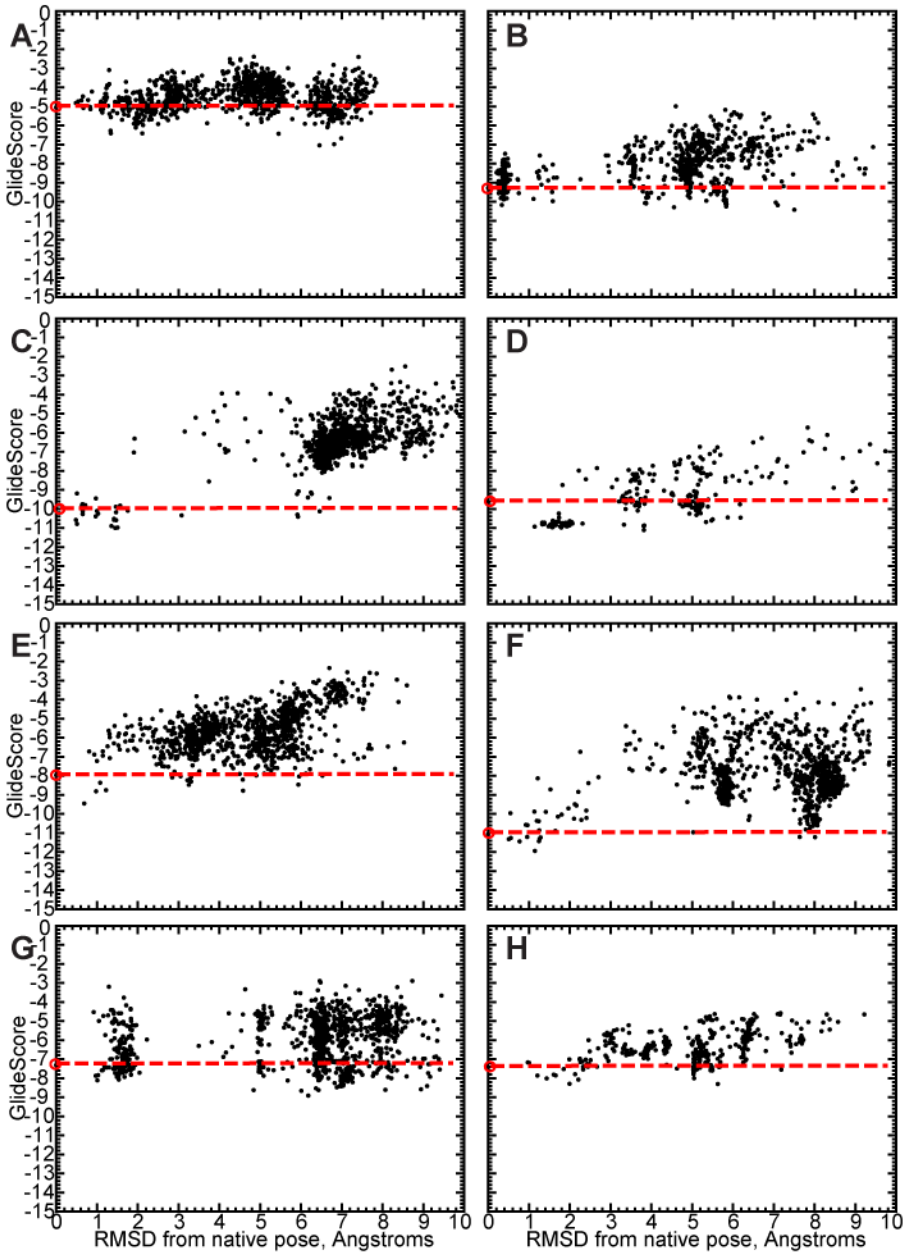

Figure S3. Root mean square deviation of protein backbone. (A) RMSD trajectory, using initial conformation as reference structure. (B) Distribution of pairwise RMSDs between simulation snapshots. Snapshots are separated by 100 time steps (~5 ns).

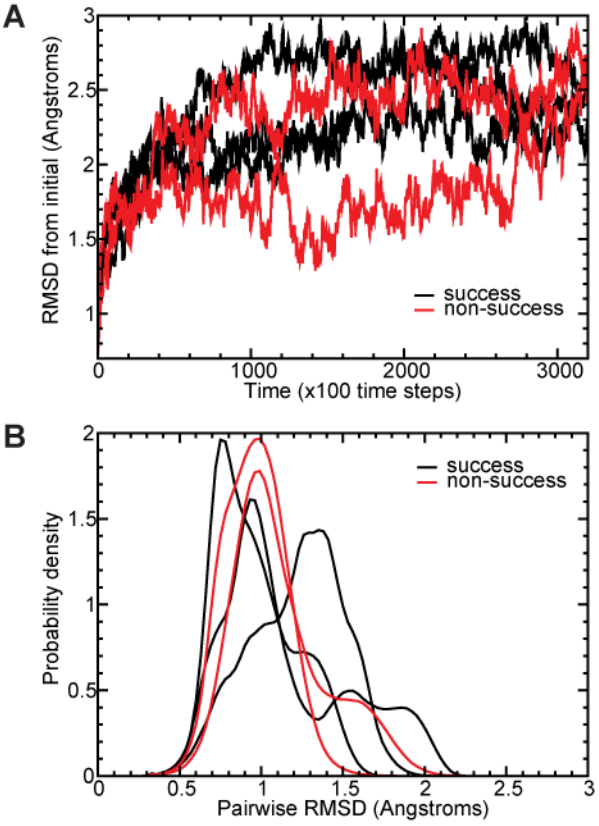

Figure S4. Root mean square deviation of pocket residues. Distributions of pairwise RMSDs between the set of pocket residues in simulation snapshots. Pocket residues are determined by the SCREEN2 server.

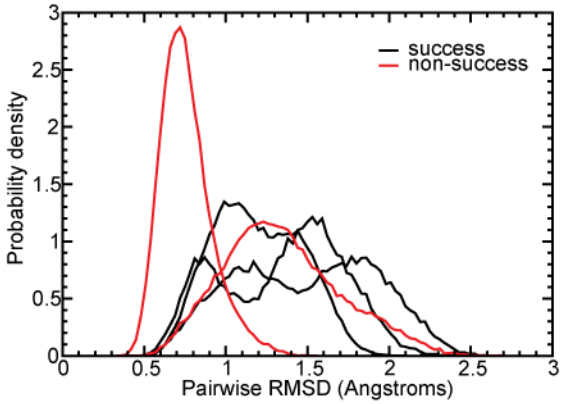

Figure S5. Ligand-target contacts. Time trajectory of ligand-target contacts. Contacts are counted between any atom of the ligand and (A) any target residue, or (B) any target atom.

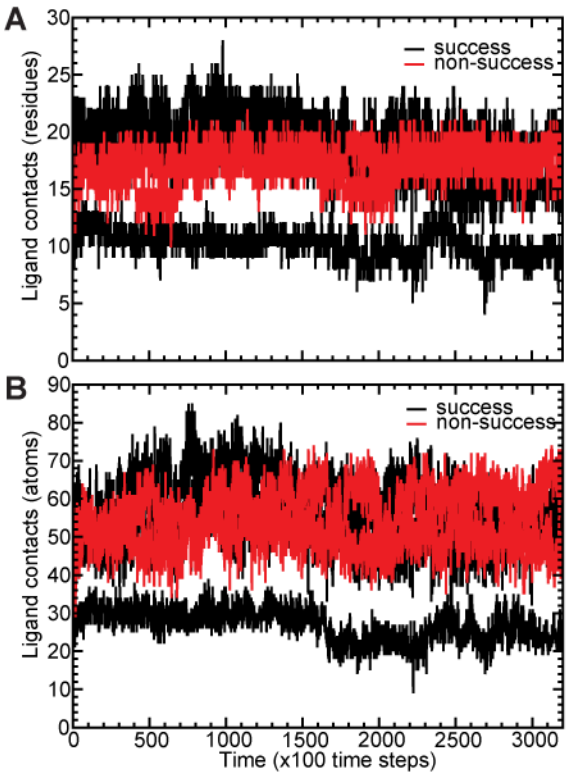

Figure S6. Correlated motion calculations. Heat maps of pairwise correlation coefficients between all residues of the target protein for (A) pantothenate synthetase (success), (B) JNK3 (success), (C) MAP kinase 14 (success), (D) CHK1 (non-success), and (E) pim-1 kinase (non-success). Positive values are red, negative values are blue, and values near zero are yellow.

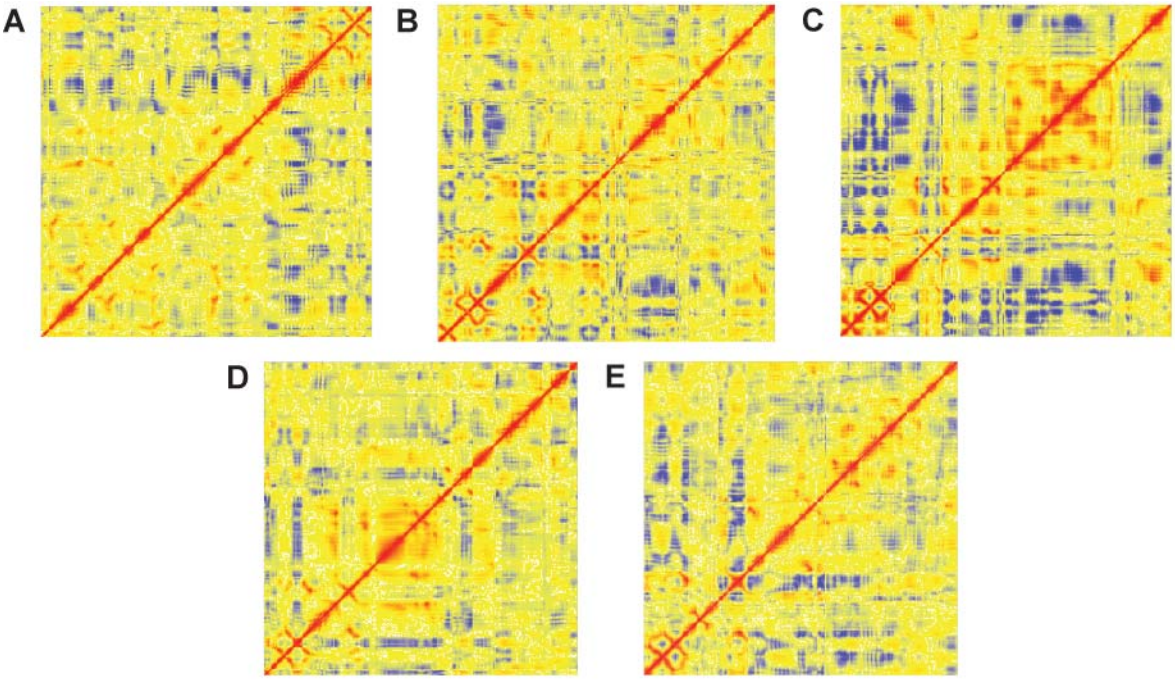

Figure S7. Representative decoy poses for acetylcholine esterase. (A) Decoy poses remaining after the application of the dual MedusaScore filter and clustering, in complex with the target. (B) MedusaScore of the remaining poses, according to their RMSD from the native pose. The red dashed line indicates the MedusaScore of the native pose. We find that the poses are structurally diverse and have varied RMSD from the native pose.

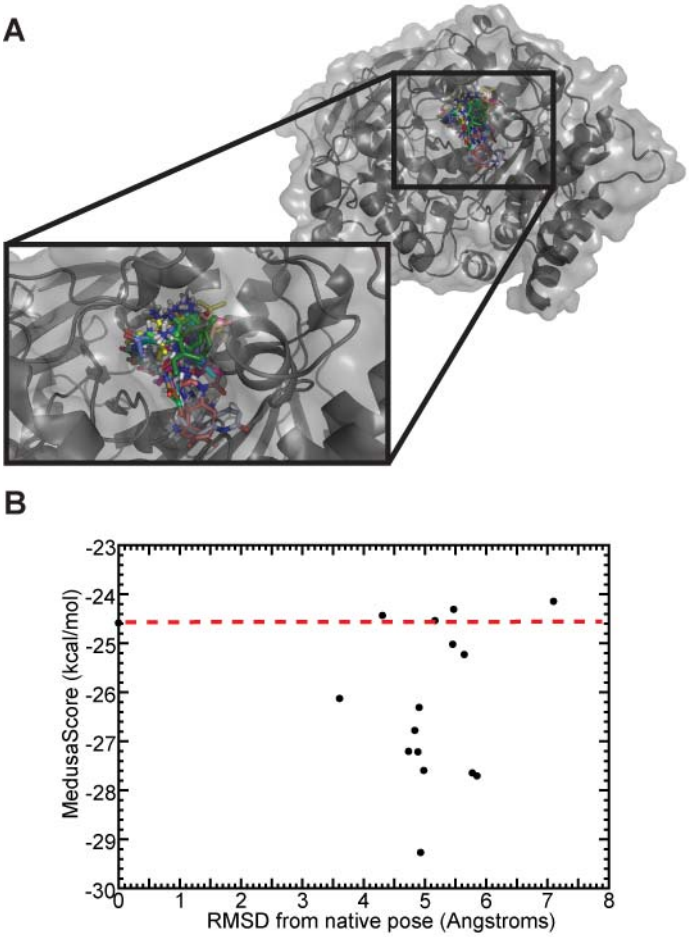

Figure S8. Representative decoy poses for C-Jun N-terminal kinase. (A) Decoy poses remaining after the application of the dual MedusaScore filter and clustering, in complex with the target. (B) MedusaScore of the remaining poses, according to their RMSD from the native pose. The red dashed line indicates the MedusaScore of the native pose. We find that the poses are structurally diverse and have varied RMSD from the native pose.

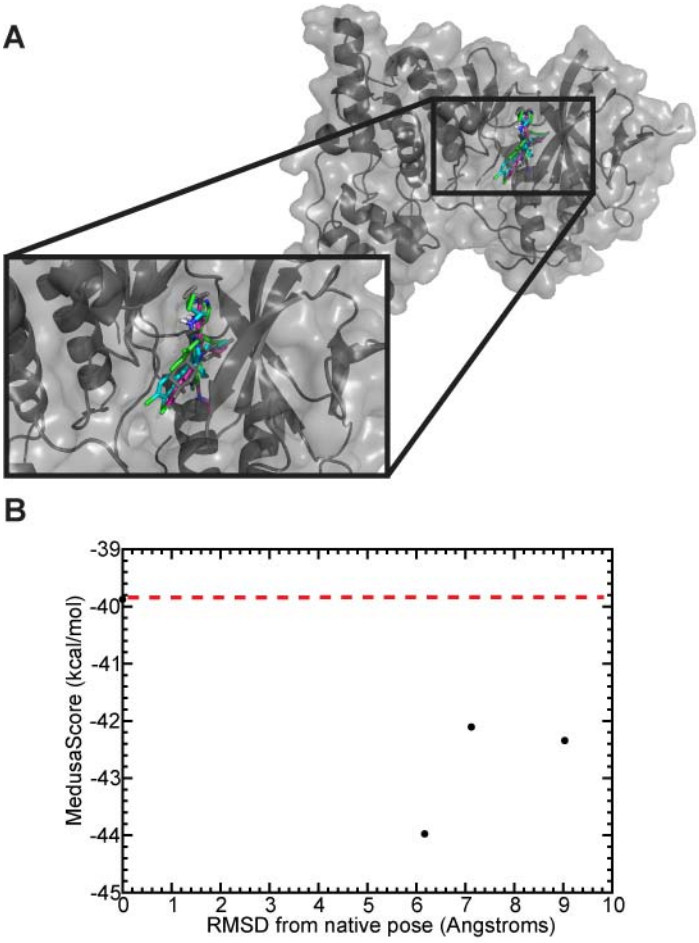

Figure S9. Representative decoy poses for LmrR. (A) Decoy poses remaining after the application of the dual MedusaScore filter and clustering, in complex with the target. (B) MedusaScore of the remaining poses, according to their RMSD from the native pose. The red dashed line indicates the MedusaScore of the native pose. We find that the poses are structurally diverse and have varied RMSD from the native pose.

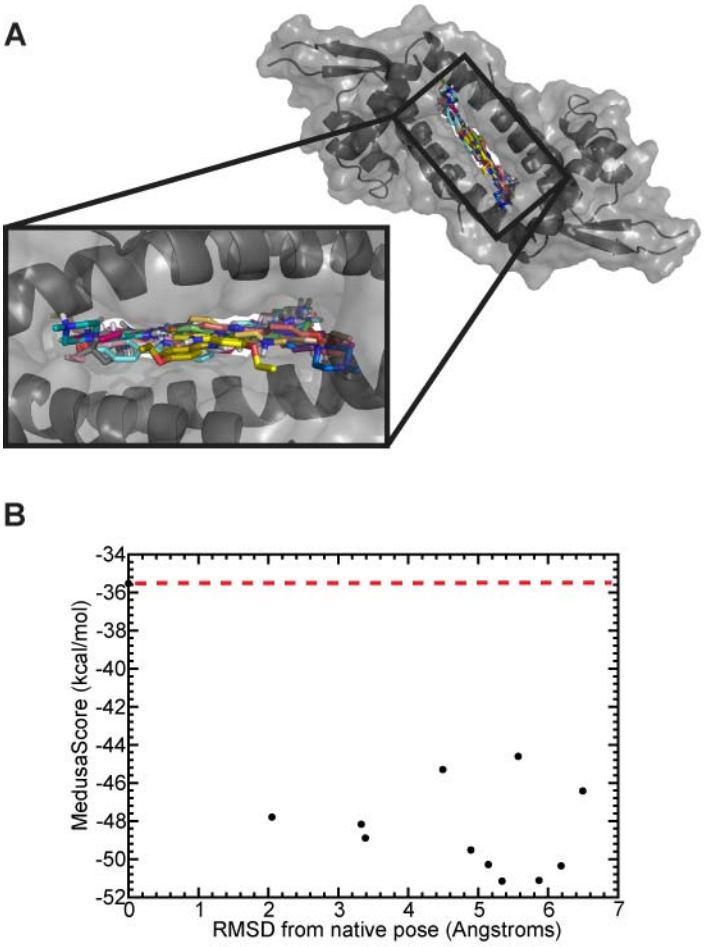

Figure S10. Representative decoy poses for tuberculosis thymidylate kinase. (A) Decoy poses remaining after the application of the dual MedusaScore filter and clustering, in complex with the target. (B) MedusaScore of the remaining poses, according to their RMSD from the native pose. The red dashed line indicates the MedusaScore of the native pose. We find that the poses are structurally diverse and have varied RMSD from the native pose.

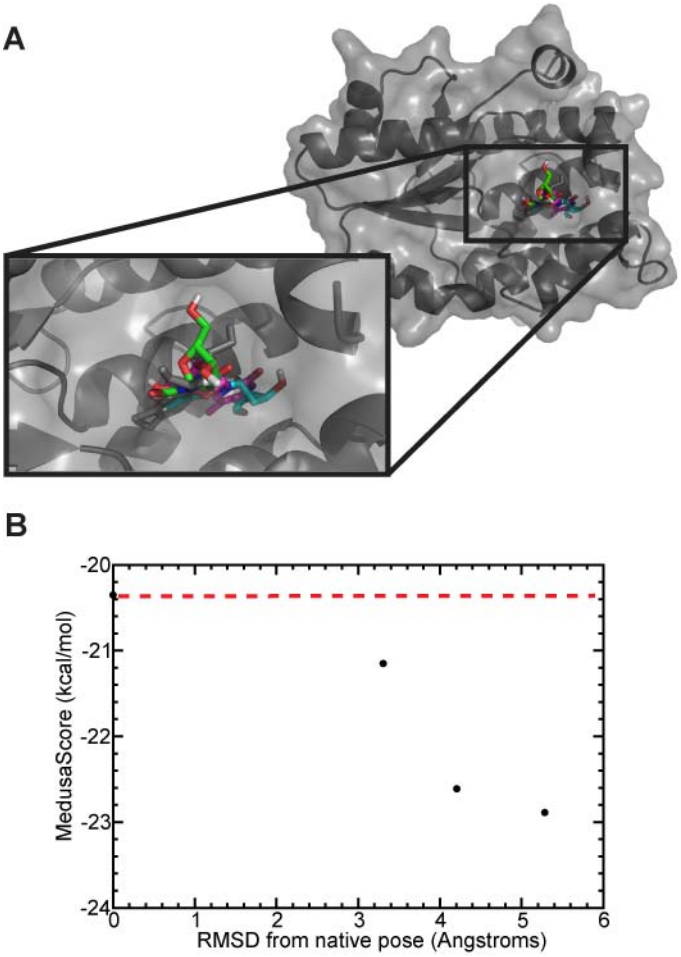

Figure S11. Representative decoy poses for MAP kinase 14. (A) Decoy poses remaining after the application of the dual MedusaScore filter and clustering, in complex with the target. (B) MedusaScore of the remaining poses, according to their RMSD from the native pose. The red dashed line indicates the MedusaScore of the native pose. We find that the poses are structurally diverse and have varied RMSD from the native pose.

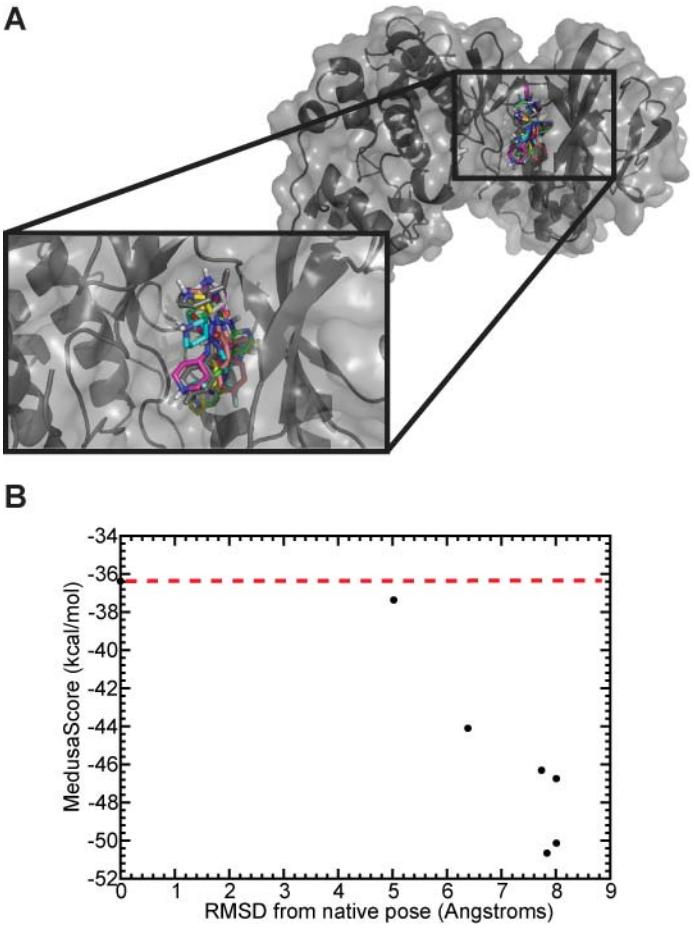

Figure S12. Representative decoy poses for Chk1. (A) Decoy poses remaining after the application of the dual MedusaScore filter and clustering, in complex with the target. (B) MedusaScore of the remaining poses, according to their RMSD from the native pose. The red dashed line indicates the MedusaScore of the native pose. We find that the poses are structurally diverse and have varied RMSD from the native pose.

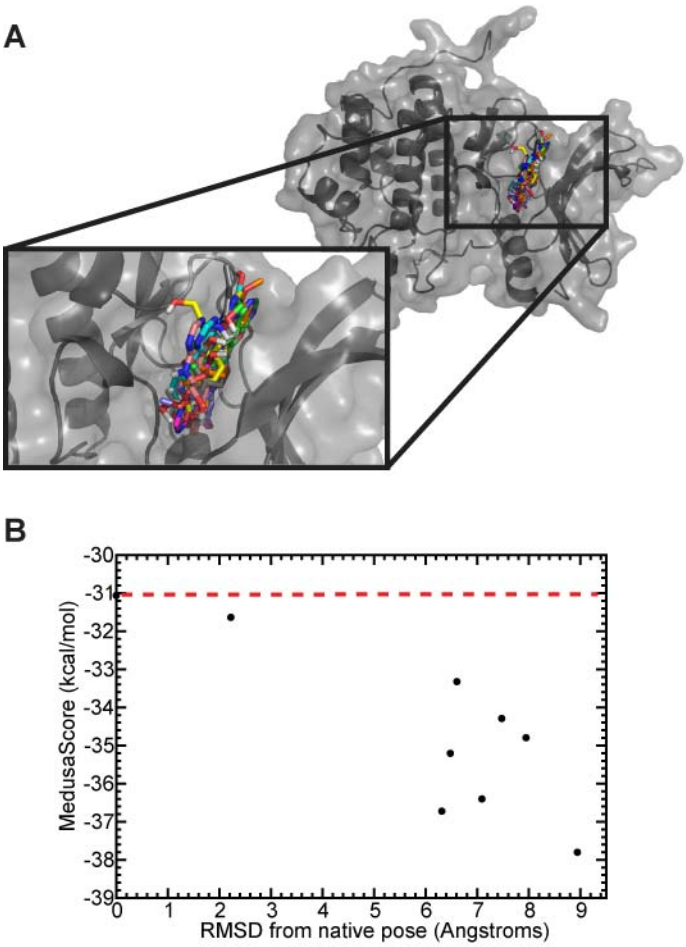

Figure S13. Representative decoy poses for pim-1 kinase. (A) Decoy poses remaining after the application of the dual MedusaScore filter and clustering, in complex with the target. (B) MedusaScore of the remaining poses, according to their RMSD from the native pose. The red dashed line indicates the MedusaScore of the native pose. We find that the poses are structurally diverse and have varied RMSD from the native pose.

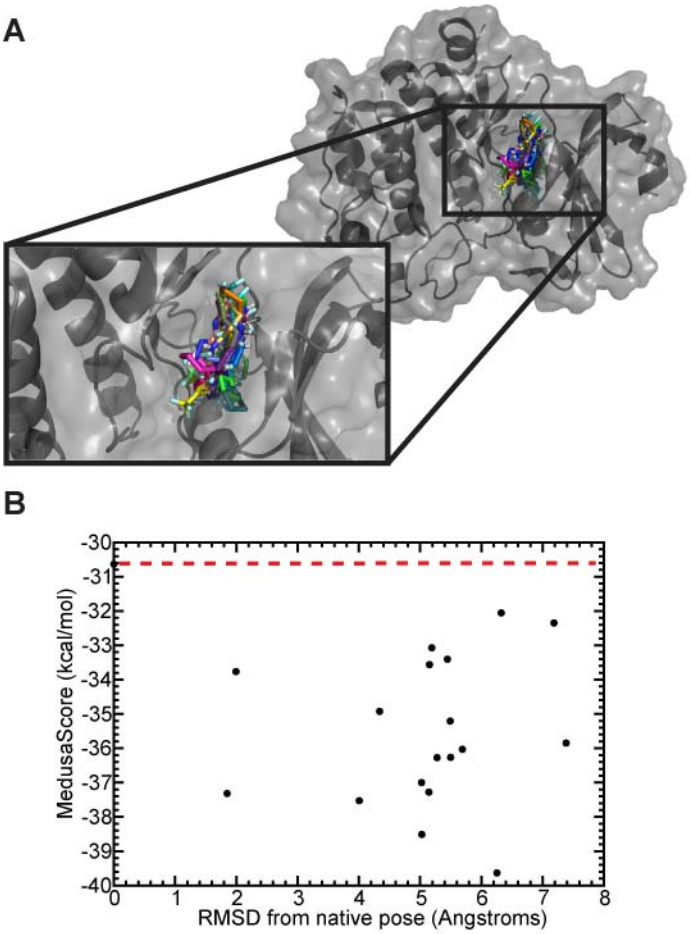

Figure S14. Decoy selection procedure. An outline of the selection process and dual MedusaScore filter that we use in our method. The dual filter (FILTER 1, FILTER 2) eliminates >98% of the docked decoys in most cases