An NMR-based scoring function improves the accuracy of binding pose predictions by docking by two orders of magnitude.

Julien Orts[a], Stefan Bartoschek[b], Christian Griesinger[c], Peter Monecke[d], Teresa Carlomagno[a*]

[a] EMBL, Structure and Computational Biology Unit, Meyerhofstrasse 1, D-69117 Heidelberg, Germany
[b] Sanofi-Aventis Deutschland GmbH, R&D LGCR/Parallel Synthesis & Natural Products, Industriepark Hoechst, Bldg. H811, D-65926 Frankfurt am Main, Germany
[c] Max Planck Institute for Biophysical Chemistry, Am Fassberg 11, D-37077 Göttingen, Germany
[d] Sanofi-Aventis Deutschland GmbH, R&D LGCR/Structure, Design & Informatics, Industriepark Hoechst, Bldg. G838, D-65926 Frankfurt am Main, Germany

* Corresponding author:

Teresa Carlomagno, EMBL, Structure and Computational Biology Unit, Meyerhofstrasse 1, D-69117 Heidelberg, Germany, Tel.: +49-6221-387 8552; Fax: +49-6221-387 8519; E.mail: teresa.carlomagno@embl.de

Supporting Information

Descriptors of similarity of relative binding modes. The agreement between the relative binding mode of L_1 and L_2 in each pair of docked complex structures with respect to the relative binding mode observed in the crystallographic structures 3DNE.pdb and 3DND.pdb is characterized by descriptors based on quaternions. RMSD calculations are impracticable due to the large number of complex pairs. Unit quaternions provide a convenient mathematical tool for representing orientations and rotations of objects in three dimensions. Compared to Euler angles they are simpler to use and bypass the problem of gimbals lock. Compared to rotation matrices, they are numerically more stable and more efficient in computation. For each transformation of a rigid body in a three-dimensional space, there exists a unique corresponding quaternion. Quaternions can be thought of as a vector, representing the axis of rotation, and an angle, representing the rotation.

Quaternions allow the unambiguous description of both the absolute binding mode of a ligand L_1 in the protein binding pocket and the relative binding mode of L_1 with respect to a reference ligand L_2. The correctness of the absolute binding mode of L_1 in a docking model can be evaluated by calculating the quaternion that describes the rotation of L_1 in the protein/L_1 complex model with respect to L_1 in the crystal structure of the protein/L_1 complex. The correctness of the relative orientation of L_1 with respect to L_2 in the complex models pair protein/L_1 and protein/L_2 is obtained by comparing the quaternions that characterize the rotations of both L_1 and L_2 in the docking models with respect to L_1 and L_2 in the crystal structures of the protein/L_1 and protein/L_2 complexes. For example, in two protein/L_1 and protein/L_2 complex models, let L_1 and L_2 be both rotated by 42° around the same axis with respect to the orientations observed in the crystal structures of the protein/L_1 and protein/L_2 complexes. In this case the absolute binding mode of both L_1 and L_2 would be incorrect, but the relative binding mode of the two ligands, namely the ligand superposition, would be correct. In other words, if one pose of $L_1^{i\in[0;4636]}$ is characterized by the same quaternion as one pose of $L_2^{i\in[0;4758]}$, the complex pair formed by combining these two poses allows the correct ligand superposition. The two quaternions describing the two binding poses L_1^i and L_2^j can be defined by five parameters $(\theta_1^i,\theta_2^j,\mathbf{q}_1^i,\mathbf{q}_2^j,s)$ where $\mathbf{q}_1^i$ and $\mathbf{q}_2^j$ represent the two axis of rotation for L_1 and L_2, respectively, θ_1^i and θ_2^j the two corresponding angles of rotation, and *s* is the shift between the center of mass of the two ligands.

We define two functions:

$$f_1 = \left(\theta_1^i - \theta_2^j\right) \cdot \left(1 - 0.7 \cdot \cos^2\left(\mathbf{q_1^i} \bullet \mathbf{q_2^j}\right)\right) \mathrm{mod}\left(360^\circ\right) \qquad \text{Eq. S1}$$

$$f_2 = \theta_{1,2}^{i,j} \cdot \left(1 - \cos^2\left(\mathbf{q_1^i} \bullet \mathbf{q_2^j}\right)\right)$$

The function f_1 allows different values of $\Delta\theta = \theta_1^i - \theta_2^j$ in dependence of the relative orientations of the axis $\mathbf{q}_1^i$ and $\mathbf{q}_2^j$. In practice we allow a larger value of $\Delta\theta$ when the two axis $\mathbf{q}_1^i$ and $\mathbf{q}_2^j$ are parallel and a lower value when the two axis are perpendicular. The function f_2 ensures that when the two axes are perpendicular $(\cos\left(\mathbf{q_1^i} \bullet \mathbf{q_2^j}\right) \approx 0)$ and $\Delta\theta \cong 0$, the rotation angles θ_1^i and θ_2^j are restricted to low values. The superposition of two binding poses of L_1 and L_2 was considered correct when:

$$f_1 \leq 29^\circ$$

$$f_2 \leq 17^\circ$$

$$s \leq 1\text{Å} \qquad \text{Eq. S2}$$

The limits of 29º and 17º were optimized empirically by visual inspection of the ligand superposition. The condition on the shift *s* ensured that the two ligands occupy the same binding pocket in the docked complexes.

Supporting Figures

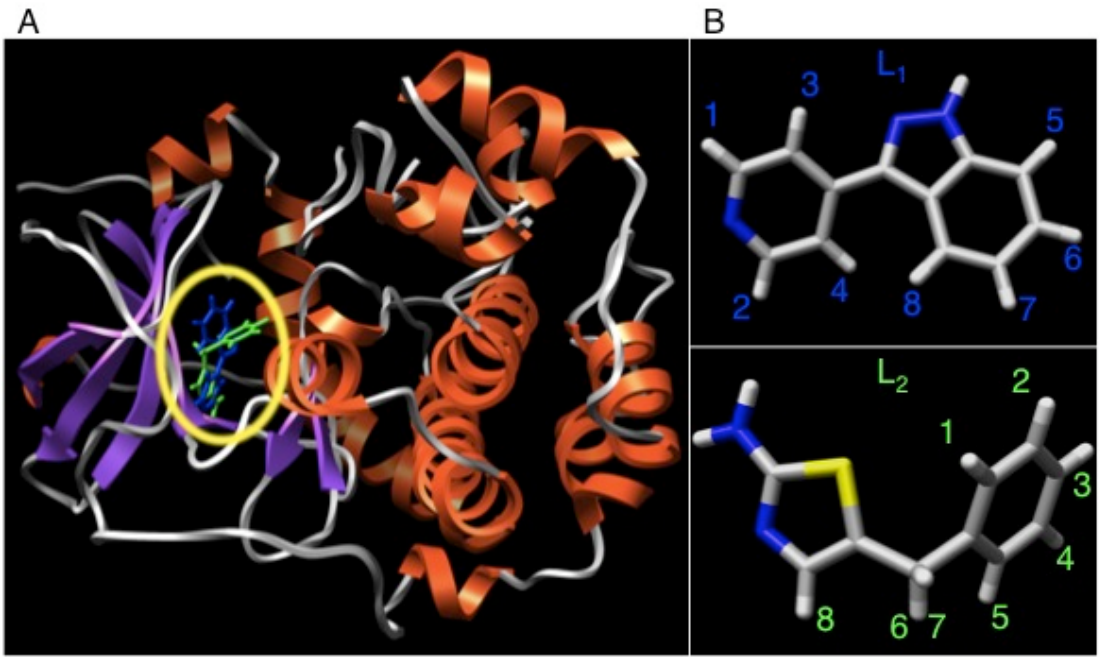

Figure S1 Superimposition of the crystal structures (panel A) of the protein kinase A (PKA) in complex with L_1 and L_2 (panel B). The crystal structures are from 3DNE.pdb and 3DND.pdb. The yellow circle indicates the ATP binding pocket, where L_1 and L_2 bind competitively.

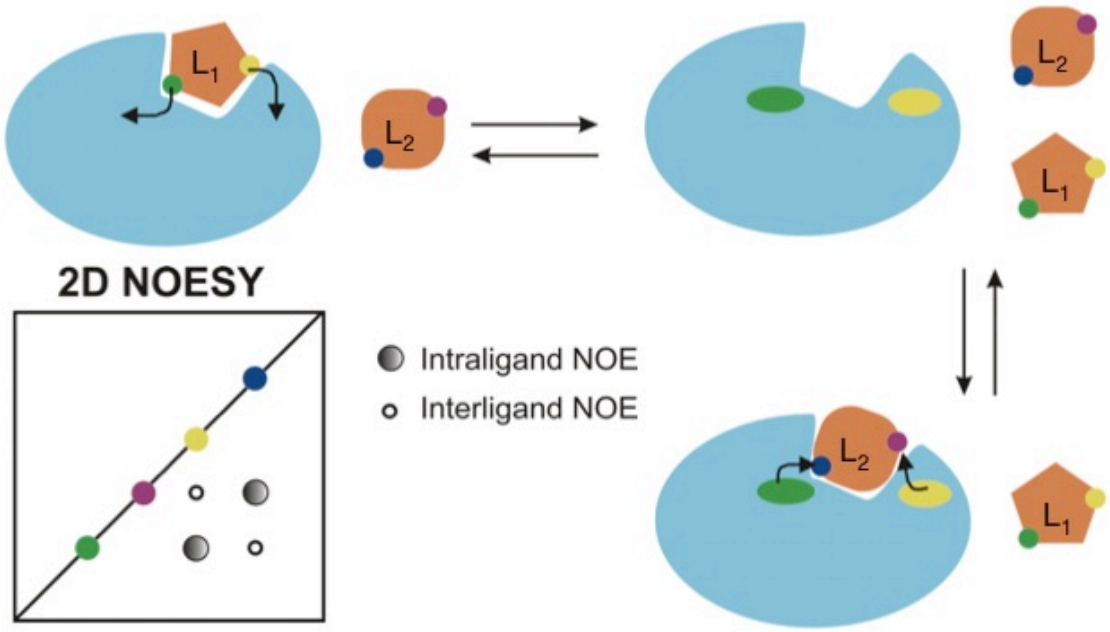

Figure S2. Schematic representation of the principle of the INPHARMA NOEs. At the beginning of the NOESY mixing time L_1 binds to the receptor and its proton H_{L1} (yellow/green) transfers magnetization to the proton of the receptor H_T (yellow/green). Being L_1 a weak binder, it dissociates from the receptor during the NOESY mixing time and leaves the binding pocket free for L_2 to occupy it. At this point the magnetization deposited on H_T by the H_{L1} proton (yellow/green) can be transferred to the H_{L2} proton of L_2 (purple/blue). This process results in a spin-diffusion mediated NOE peak between H_{L1} (yellow/green) and H_{L2} (purple/blue).

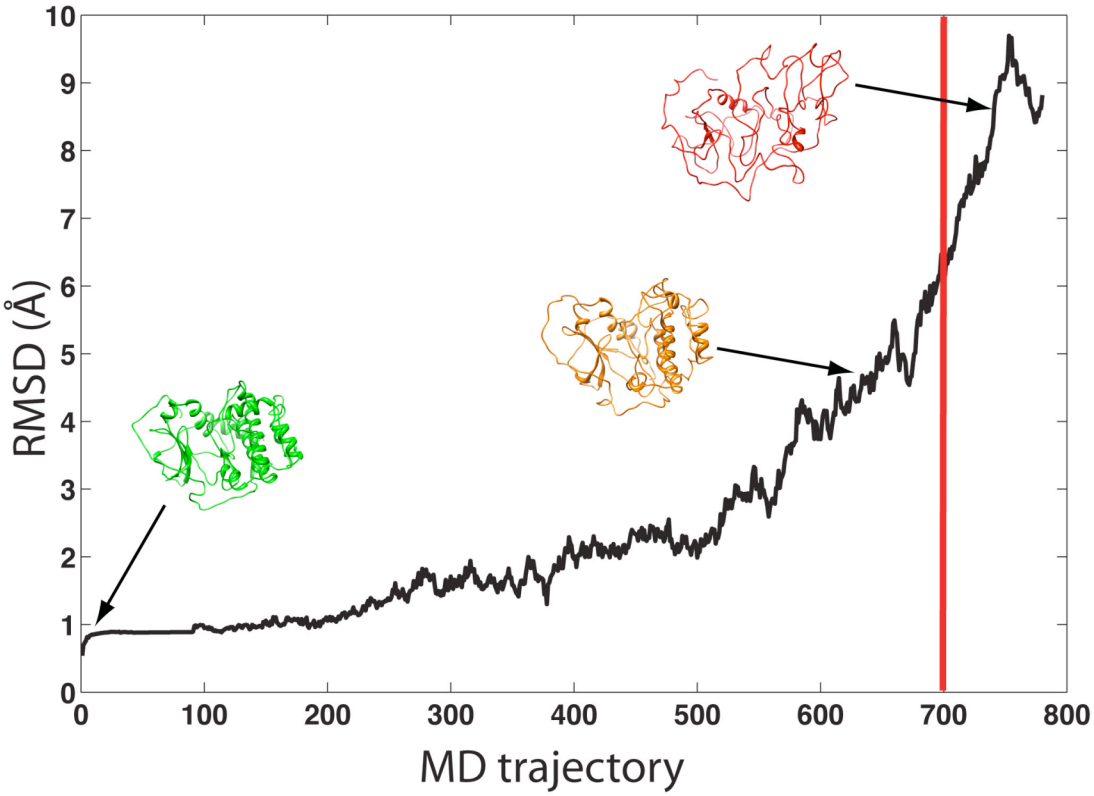

Figure S3. RMSD values of the ligand binding pocket of the protein PKA from the crystal structure of 3DNE.pdb along the MD trajectory. The x-axis represents the MD trajectory and each unit corresponds to one snapshot (780 total snapshots). The vertical red line delimitates the protein models retained for docking (~ 700). Structures beyond this line are discarded as the protein starts unfolding. Three structures are depicted on the graph to show the evolution of the protein along the trajectory.

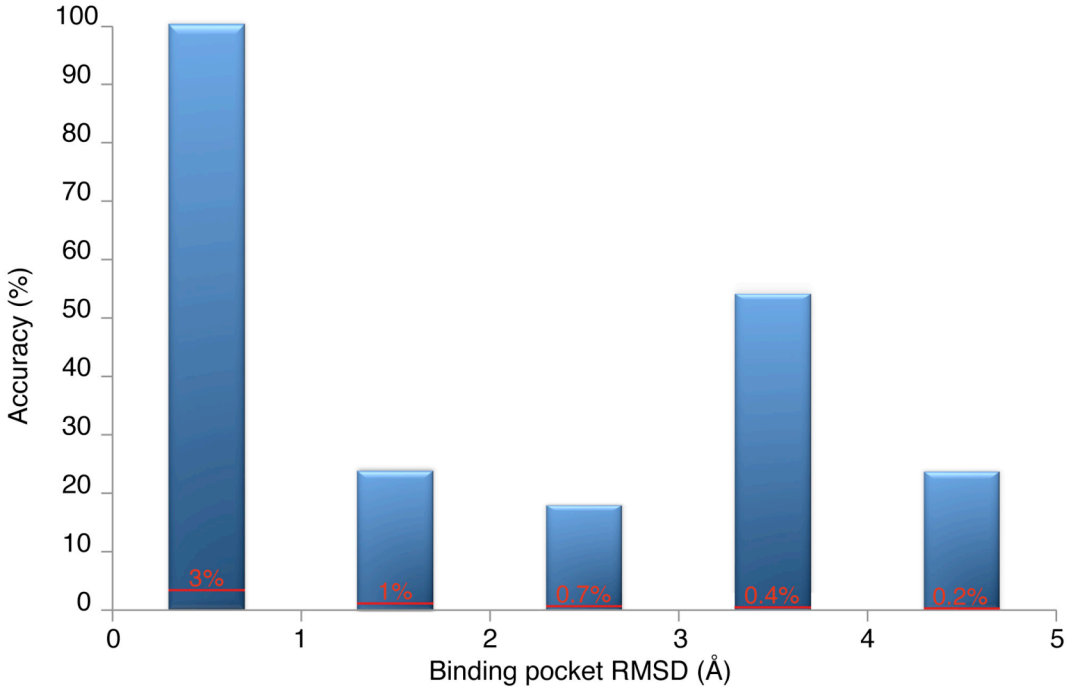

Figure S4. Accuracy of the INPHARMA predictions as a function of the quality of the receptor structure. The x-axis represents the (protein only) binding pocket RMSD of the receptor models used in the docking from the crystallographic structure of PKA in the complexes PKA/L_1 (3DNE pdb). The accuracy on the y axis is defined as the number of complex pairs reproducing the correct ligands superposition (relative binding mode of L_1 and L_2) divided by the total number of pairs selected by INPHARMA. The numbers over each bar in red represent the accuracy before applying the INPHARMA score. In this case the accuracy is the number of the complex pairs showing the correct ligands superposition divided by the total number of complex pairs selected by the energy function of the docking program. The docking for this dataset was performed by GLIDE.

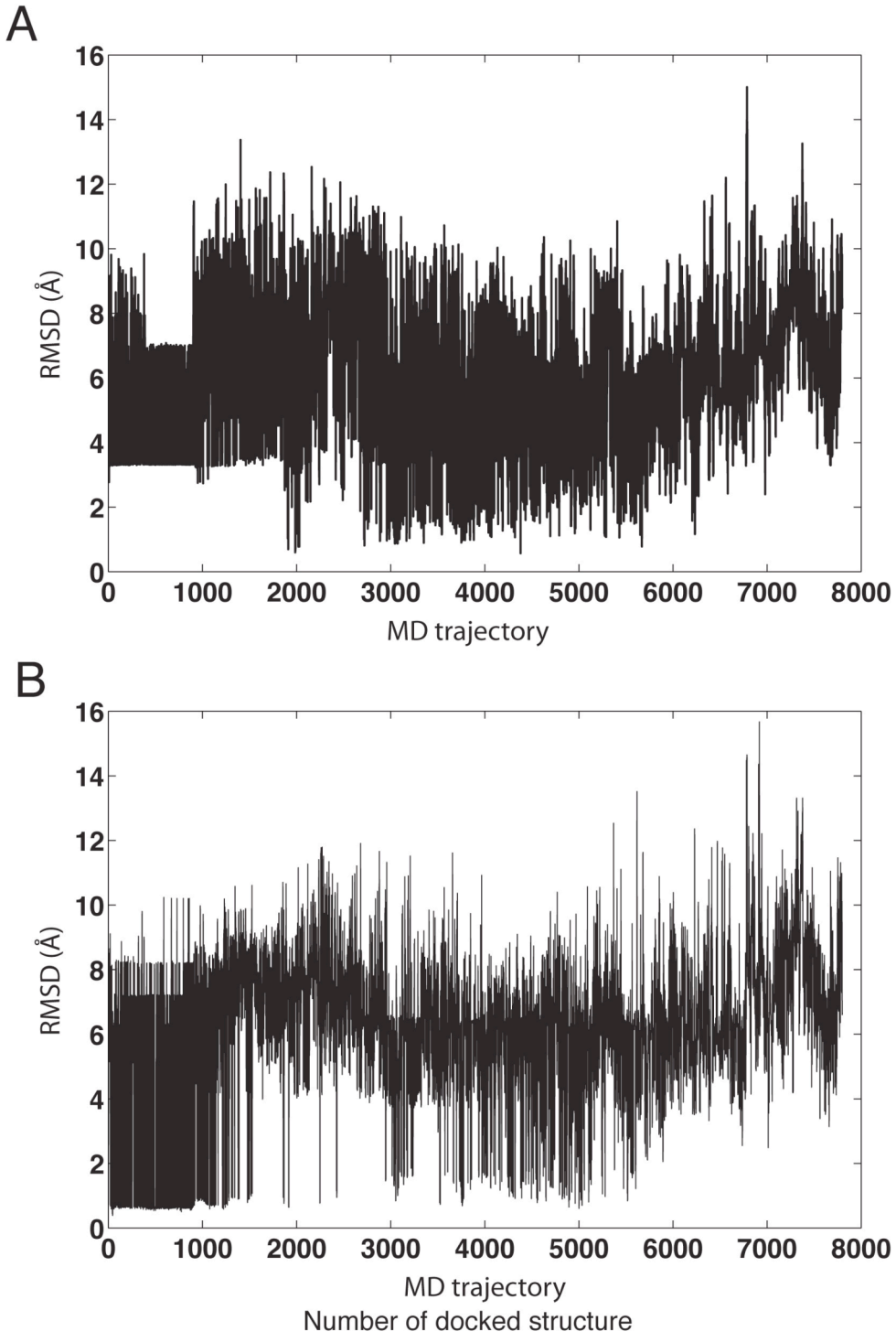

Figure S5. Representation of the quality of the docking poses measured by the ligand RMSD to the native binding mode found in the crystal structures (1DNE.pdb and 1DND.pdb) for L_1 (A) and L_2 (B). The x-axis represents the docked binding poses generated by SURFLEX on each receptor model resulting from the free-protein MD simulation. The best 10 poses for each protein model were kept during docking, resulting in a total of 7800 complex structures for each ligand. Note that only the first 7000 complex structures were retained, corresponding to the first 700

proteins models in the MD simulation. The docking program has difficulties in docking L_1 in the correct orientation (RMSD > 3 Å), especially when docking to the first 200 protein structural models (corresponding to the first 2000 docking poses).

Table S1. INPHARMA NOE intensities, $I_{INPHARMA}$, were normalized to the diagonal peak with equal ω_2 frequency at the lowest mixing time. The experimental conditions are described in the Experimental Section.

ω_2	ω_1	$I_{INPHARMA}$, 800 MHz	
L_1	L_2	τ_m = 300 ms	τ_m = 600 ms
H8	H1,H3,H5	none	0.0096
H3,H4	H1,H3,H5	0,0026	0.0067
H1,H2	H1,H3,H5	0,0016	0.0055
H3,H4	H6,H7	0,0016	0.0061
H5	H6,H7	none	0.0048
H5	H1,H3,H5	0,0029	0.0080
H1,H2	H6,H7	0,0011	0.0028

Additional weak NOEs were measured at 900 MHz and τ_m = 600 ms. These NOEs were not used in a quantitative way, that is they did not enter the calculation of the Pearson correlation coefficient. Instead, poses were retained that showed the same ranking in magnitude as the experimental data, namely:

1. $I_{INPHARMA}$ (H5(L_1), H8(L_2)) + $I_{INPHARMA}$ (H8(L_1), H8(L_2)) < $I_{INPHARMA}$ (H1,H2(L_1), H8(L_2)) + $I_{INPHARMA}$ (H3,H4(L_1), H8(L_2))
2. $I_{INPHARMA}$ (H5(L_1), H6,H7(L_2)) < $I_{INPHARMA}$ (H8(L_1), H6,H7(L_2))