Supporting Information

Combining structural and coevolution information to unveil allosteric sites

Giuseppina La Sala[1,#,*], Christopher Pfleger[2,#], Helena Käck[3], Lisa Wissler[3], Philip Nevin[4], Kerstin Böhm,[4] Jon Paul Janet[1], Marianne Schimpl[5], Christopher J. Stubbs[5], Marco De Vivo[6], Christian Tyrchan[7], Anders Hogner[1], Holger Gohlke[2,8,*], Andrey I. Frolov[1,*]

1. Medicinal Chemistry, Research and Early Development, Cardiovascular, Renal and Metabolism (CVRM), BioPharmaceuticals R&D, AstraZeneca, Gothenburg, Sweden.

2. Mathematisch-Naturwissenschaftliche Fakultät, Institut für Pharmazeutische und Medizinische Chemie, Heinrich-Heine-Universität Düsseldorf, 40225 Düsseldorf, Germany.

3. Mechanistic and Structural Biology, Discovery Sciences, BioPharmaceuticals R&D, AstraZeneca, Gothenburg, Sweden.

4. Discovery Biology, Discovery Sciences, BioPharmaceuticals R&D, AstraZeneca, Gothenburg, Sweden.

5. Mechanistic and Structural Biology, Discovery Sciences, BioPharmaceuticals R&D, AstraZeneca, Cambridge, UK.

6. Laboratory of Molecular Modeling and Drug Design, Istituto Italiano di Tecnologia, Via Morego 30, 16163 Genoa, Italy.

7. Medicinal Chemistry, Research and Early Development, Respiratory & Immunology (R&I), BioPharmaceuticals R&D, AstraZeneca, Gothenburg, Sweden.

8. John von Neumann Institute for Computing (NIC), Jülich Supercomputing Centre (JSC), Institute of Biological Information Processing (IBI-7: Structural Biochemistry), and Institute of Bio- and Geosciences (IBG-4: Bioinformatics), Forschungszentrum Jülich GmbH, 52425 Jülich, Germany.

# These authors contributed equally.

* Corresponding authors.

Table of Contents

Additional text

Computational Methods

System preparation and MD simulations. MD simulations of Glucocorticoid Receptor (GR), Lymphocyte Function-associated Antigen 1 (LFA-1), Branched Chain Ketoacid Dehydrogenase Kinase (BCKDK), Mitogen-Activated Protein Kinase 14 (MAPK14 or p38-α), and Methionine Adenosyl Transferase 2A (MAT2A) were performed both in the *apo* (i.e., unbound) and in the *holo* (i.e., bound orthosteric site ligand) forms. For MAT2A, we have additionally simulated the protein in complex with compound 1. MD simulations of *apo* and *holo* systems were used for pocket detection and rigidity analysis by CNA, respectively. The X-ray structures of the protein in the unbound form (PDB ids 3F74 chain C for LFA-1, 5ETC for p38-α and 6FAJ for MAT2A, 1GKX for BCKDK) were used as starting structures for the *apo* systems. If unbound states were not present in the Protein Data Bank (PDB), we chose the same *holo* X-ray structure but removed the orthosteric ligand (PDB id 4UDC for GR). As starting structures for the *holo* systems, we used PDB ids 1GKZ for BCKDK, 3TCX for LFA-1, 3S3I for p38-α, 6FBP and 8OOG for MAT2A in complex with the orthosteric ligand and compound 1, respectively.

We used the Protein Preparation Wizard workflow implemented in Maestro to add missing hydrogens and model missing loops. The Amber ff14SB force field for proteins and the GAFF force field for ligands after computing the point charges with the RESP method at the HF/6-31G* level of theory were employed. Each system was solvated in a TIP3P[1] water box using a buffer distance of 12 Å between the solute and the edge of the box. Either Na$^+$ or Cl$^-$ ions were used to neutralize the total charge of the system. After a minimization step, each system was submitted to a 500 ps-long thermalization phase to reach a target temperature of 310 K and a pressure of 1 bar. First, three steps of 50 ps-long MD simulations in the NVT ensemble were performed to gradually heat the system to 310 K. Each of these steps was followed by 50 ps-long MD simulations in the NPT ensemble. Harmonic restraints were applied to the backbone atoms with a spring constant of 1000 kJ/(mol·Å^2). Then, the restraints were gradually released in 200 ps-long MD simulations in the NPT ensemble. A Parrinello–Rahman barostat[2] and a velocity-rescaling thermostat[3] were employed. The LINCS algorithm[4] was used to constrain covalent bonds involving hydrogen atoms. We used a short-range neighbor list cutoff of 10 Å, while long-range electrostatics were treated using the particle mesh Ewald (PME) method[5] with a Fourier grid spacing of 1.2 Å and cubic spline interpolation. Periodic boundary conditions were applied, and a time step of 2 fs was set for integrating the equations of motion.

We performed 500 ns of production phase in the NPT ensemble for each apo system (MD trajectories were used for MDpocket analysis). Ten replicas of 50 ns each were performed for the holo systems (MD trajectories were used for CNA). All of the simulations in this study were performed with the Gromacs 2016.4 engine.[6]

Pocket detection. Fpocket[7] with default settings was used to identify pockets in the *apo* X-ray structures. Pockets formed in MD trajectories were detected with MDpocket[8] using the workflow depicted in **Figure S11**. In detail, each MD trajectory was aligned on the protein's backbone and submitted to MDpocket. As output, we obtain a map of the density of alpha spheres in each pocket and a frequency map corresponding to the pockets' opening frequency (see ref [8] for more details) (**Figure S11** a). The density map can be represented at high and low isocontours, and the optimal isovalue for result visualization depends on the protein's size and surface features.[8] For LFA-1, we used isocontours at isovalues=1, while for the remaining systems, we employed isovalues=2. These isovalues correspond to an opening frequency of 20 and 30%, respectively. We applied a DBSCAN clustering algorithm to separate the 3D points of the density map into separate pockets (**Figure S11** b). Each pocket is then defined as the residues surrounding the 3D points (**Figure S11** c). The detected pockets were filtered by removing: i) the orthosteric pockets; ii) small pockets surrounded by less than five residues; iii) inaccessible and buried pockets identified via visual inspection (**Figure S11** d) (see Supplementary file).

Druggability Score For each pocket, we estimated the druggability score (DS) according to the model implemented by Schmidtke *et* al.[9] The DS combines the mean local hydrophobic density, hydrophobicity score, and

polarity score descriptors. We obtained the three descriptors by re-running MDpocket on each pocket identified in the initial run (**Figure S11** e). The descriptors' values were averaged over the MD trajectories and fed to the model published in ref [9] to obtain the final DS (**Figure S11** e).

Relationship between residue conservation and allosteric pockets. In this work, we used the residue co-evolution calculated via SCA[10, 11, 12] to highlight pockets that are more likely to be allosteric amongst the protein's pocket ensemble. We also evaluated whether using the information from residue conservation might have given better or comparable results. The conservation of each amino acid a at position i is measured from the deviation of the observed distribution of amino acids at this position from a background distribution expected by natural drift (Eq. 1):

$$D_i^a = f_i^a \ln\frac{f_i^a}{q^a} + (1 - f_i^a) \ln\frac{1 - f_i^a}{1 - q^a} \qquad (1),$$

f_i^a is the observed frequency of the amino acid and q^a is the background expectation. D_i^a is the Kullback-Leibler relative entropy.[13]

For each pocket, we calculated the average entropy and the median entropy obtained from the pocket residues (see **Figure S11**). Also, we report the maximum entropy value observed in the pocket. Values are reported in **Table S3.** In three out of five systems, the highest entropy value is found in the orthosteric pockets, indicating that highly conserved residues are found in the active site, underscoring their relevance to protein catalytic activity. Except for GR, allosteric pockets do not have high residue conservation. Although the dataset is limited, these results suggest that there is not a clear link between residue conservation and allostery, in agreement with the work of Panjkovich *et al.*[14]. Thus, we decided to employ residue co-evolution as a method to investigate allosteric pockets.

Statistical Coupling Analysis. We performed a statistical coupling analysis of the LFA-1, GR, BCKDK, p38-α, and MAT2A proteins using the approach developed by the Ranganathan group[15,10] and adopted in similar studies.[16] First, we generated the multiple sequence alignment (MSA) for each protein. To do so, we adopted HMMER 3.3[17] to search the UniRef50 database for sequences similar to LFA-1 (UniProt P20701), GR (UniProt P04150), BCKDK (UniProt Q00972), p38-α (UniProt Q16539), and MAT2A (UniProt P31153) with e-score < 0.0001. All sequences were aligned with ClustalW2[17] using default settings. We removed the sequences with an identity < 10% to the reference sequence and those with a fraction of gaps > 15% with respect to the reference sequence's length. We estimated the conservation of different positions for each MSA by using the Kullback-Leibler relative entropy (see eq 1). Next, we calculated the conservation-weighted positional correlation matrix as described in the work of Halabi *et al.*[15]. The correlation matrix was submitted to the eigenvalue decomposition procedure to retrieve the most statistically significant eigenvalues (k^*). To obtain k^*, we compared the spectral decomposition of the actual alignment with that of many trials of randomized alignments in which the amino acids at each position are scrambled independently (k^*=8 for LFA-1, k^*=6 for GR, k^*=9 for BCKDK, k^*=11 for p38-α, k^*=10 for MAT2A). The independent component analysis (ICA) transforms the top k^* eigenmodes into maximally independent components. We used SCA toolbox 5.0 implemented in MATLAB for these computations.

Most points are grouped near the origin when plotting the positions along the orthonormal axes (the ICs) (black dots in **Figure S12**). These points represent weakly correlated positions and, therefore, will be discarded from the subsequent analysis. The remaining points are grouped along separate orthonormal axes and correspond to evolutionarily correlated positions (cyan dots in **Figure S12**). The cutoff that divides the two groups is set automatically for each system and is obtained to leave out ~85% of the data points in the origin of the orthonormal axes (**Figure S12**).

In the original implementation, the evolutionarily correlated positions (see cyan dots in **Figure S12**) can be further clustered to form sectors, which are independent residues having a distinct functional role in the protein.[15] However, the separation of different sectors requires manual intervention and deep knowledge of the system, which

discourages the application of automatic methods for sector identification. We avoided the sector analysis since we aimed to exclude manual intervention from our protocol. We were aware that all the evolutionarily correlated positions examined here might be relevant for various protein functions, and not be linked necessarily with the formation of allosteric pockets. This is a limitation of the SCA methodology that we attempted to mitigate by integrating it with CNA as a second approach for allosteric signal investigation.

Finally, we assigned the Coverage Score (CS) to each pocket, which corresponded to the percentage of evolutionarily correlated amino acids in a specific pocket identified in the MD simulations. Also, we examined the importance of residues conservation for the identification of functional pockets.

Constraint Network Analysis. Rigidity analyses were performed using the CNA software package.[18] We applied CNA on ensembles of network topologies generated from conformational ensembles provided as input. Average stability characteristics were calculated by constraint counting on each topology in the ensemble. The FIRST software (version 6.2)[19] was used to construct each network of covalent and non-covalent (hydrophobic tethers, hydrogen bonds and salt bridges) constraints. The hydrogen bond energy E_{HB} is determined from an empirical energy function.[20] Hydrophobic tethers between carbon and sulfur atoms were considered if the distance between these atoms was less than the sum of their van der Waals radii (C: 1.7 Å, S: 1.8 Å) plus 0.25 Å (D_{cut}).[21] Biomolecules generally display a hierarchy of structural stability, reflecting the modularity of their structure.[22] To identify this hierarchy, a "constraint dilution trajectory" of network states $\{\sigma\}$ is analyzed by successively removing non-covalent constraints from an initial network topology.[21,22,23,24,25] Here, hydrogen bond constraints (including salt bridges) are removed by their increasing strength. That way, for network state σ, only hydrogen bonds are retained that have an energy $E_{HB} \leq E_{cut}(\sigma)$.

Allosteric signaling through constraint networks. We used a per-residue decomposition scheme to identify the extent to which single residues contribute to the allosteric signaling. First, neighbor stability maps ($rc_{ij,\text{neighbor}}$) are derived from the "constraint dilution trajectory"; they contain information accumulated over all network states $\{\sigma\}$ along the trajectory.[26, 27] More precisely, stability maps monitor the persistence of rigid contacts for pairs of residues during a bond dilution process; a rigid contact is present as long as two residues belong to the same rigid cluster c of the set of rigid clusters C^{Ecut}. As our focus is on short-range rigid contacts, only pairs of heavy atoms of the residue pair $R_{\{i, j\}}$, $A_{\{k \in i, l \in j\}}$, separated by a distance $d \leq 4.5$ Å are considered. The resulting neighbor stability map (Eq. 2)[27] relates to the local stability in the network.

$$rc_{ij,neighbor} = \min\left(E_{cut}\middle|\exists c \in C^{Ecut}\colon\left(R_i \wedge R_j \in c\right) \wedge \left(d\left(A_{\{\exists k \in i, \exists l \in j\}}\right) \leq 4.5\text{Å}\right)\right) \tag{2},$$

A chemical potential energy $E_{i,\text{CNA}}$ of residue i is obtained by summation over all n short-range rigid contacts in which the residue is involved (Eq. 3)

$$E_{i,CNA} = \frac{1}{2}\sum_{i \neq j}^{n} rc_{ij,neighbor} \tag{3},$$

which is then used in a linear response approximation to obtain the per-residue decomposition (Eq. 4)

$$\Delta G_{i,CNA} = \alpha\left(\langle E_{i,CNA}^{perturbed}\rangle - \langle E_{i,CNA}^{ground}\rangle\right) \tag{4},$$

with α being 0.5.[26] Based on the computed $\Delta G_{i,CNA}$ values from Eq. 4, we performed ROC analyses to evaluate whether the pockets identified in MD simulations comprise more residues with larger $\Delta G_{i,CNA}$ values than other surface residues (see definition of surface residues below). To quantify the enrichment of larger $\Delta G_{i,CNA}$ values within the pockets, we computed the area under the curve (AUC)

Definition of surface residues. To find the surface-exposed residues for each system, we used the PyMol script *findSurfaceResidues.py* (see http://pymolwiki.org/index.php/FindSurfaceResidues). Surface exposed residues require a minimal surface area of ≥ 3.5 Å^2 and have been identified based on the closest to average structure from the *apo* MD simulations (see above). Finally, we built the set difference to obtain only those surfaces residues ($R_{surf,no}$) that are not part of the orthosteric site (R_{ortho}) (Eq. 5)

$$R_{surf,no} = R_{surf} - R_{ortho}$$ (5).

Three-parameter ranking model. The pockets identified in the MD simulations were ranked according to the three-parameter model that considers the druggability, the residue co-evolution, and the rigidity analysis results. In brief, we assigned a value (V) to each pocket according to Eq 6:

$$V = DS + CS \times k + AUC$$ (6).

where DS is the druggability score, CS is the coverage score from SCA, which is multiplied by a coefficient k (k = 1.7) and AUC is the value from the rigidity analysis.

We introduced k for SCA because we noticed that for all the detected pockets both the CNA score (**Figure S13A**, red line) and the Druggability Score (blue line) span from 0 to 1, while the SCA score (grey line) has a distribution that is systematically below the other two scores, ranging from 0 to ~0.6. Therefore, we decided to scale the SCA contribution to increase its relative weight in the final score. We have computed the final score of the allosteric pockets for the five systems varying k from 1.0 to 3.0. The best performance was achieved using k between 1.6 to 1.9 (see **Figure S13B**). However, we decided to adopt k=1.7 such that the SCA score spans from 0 to 1 (see black line in **Figure S13A**), similarly as the CNA and the Druggability Score. In each system, the identified pockets were ranked according to V. We hypothesized that pockets in the top 33% of the ranking ($R_{33\%}$) have a high likelihood of being allosteric.


Experimental Methods

Expression and purification of human MAT2A. For crystallization, full-length human MAT2A was recombinantly expressed in E. coli with an N-terminal cleavable hexahistidine tag. The expression plasmid was a gift from Nicola Burgess-Brown at the Structural Genomics Consortium. For surface plasmon resonance studies, an expression construct His6-TEV-Avi-MAT2a (2-395) was generated. MAT2A was purified from clarified lysate by Ni2+-immobilised metal affinity chromatography (IMAC), proteolytic cleavage of the hexahistidine tag, and size exclusion chromatography. In vitro biotinylation was carried out on the Avi-tagged protein using the BirA biotinylation kit from Avidity according to the manufacturer's instructions. Excess reagents were removed by passing over a desalting column. The protein was snap-frozen and stored at -80°C.

X-ray analysis of the MAT2A / compound 1 complex. Full-length human MAT2A at a concentration of 20 mg/ml (0.47 mM) in a buffer containing 10 mM Hepes pH 7.5, 500 mM NaCl, 5% glycerol, and 0.5 mM TCEP was supplemented with a 1.5-fold molar excess of S-adenosylmethionine. Sitting drop vapor diffusion crystallization experiments with 150 nl protein and 150 nl of a reservoir solution containing 8 % PEG 8000, 12 % ethylene glycol, 0.1 M Hepes pH 8.0 gave bar shaped crystals within 12 hours. Crystallisation reliability benefitted from micro-seeding. Crystals were soaked with 30 mM compound (final DMSO concentration of 30 %) for 16 h, frozen without further cryoprotection, and X-ray diffraction data were collected at Soleil beamline Proxima 2. Images were processed with autoproc, and additional programs from the CCP4 suite.[28] The structure was solved by molecular replacement using AMoRe[29] and refined with Buster[30] and Coot.[31] Ligand restraint dictionaries were generated with Grade.[32] Data

collection and refinement statistics are reported in **Table S7**; atomic coordinates and structure factors are available from the Protein Data Bank (accession code 8OOG).

SPR assay for MAT2A. SPR experiments were run at 25 °C using a Biacore T200 instrument (Cytiva). A Series S Sensor Chip SA (Cytiva) was docked into 20 mM Bicine-Na pH 7.5, 100 mM KCl, 5 mM MgCl2, 0.1 mM TCEP, 0.05% (w/v) Tween-20, 2% (v/v) DMSO and the chip conditioned with 3 x 60 s injections of 50 mM NaOH/1 M NaCl. Biotinylated Avi-tagged MAT2A was immobilised onto the surface by injecting for 20 min (10 μg/mL Avi-MAT2A, 20 min injection at 2 μL/min) achieving an immobilisation density of ~6500 RU, after which remaining biotin binding sites were blocked (100 μM amine-PEG2-biotin (ThermoScientific), 60 s at 10 μL/min). Compound 1 was tested as a 10-point 1.5-fold serial dilution series from 667 μM top concentration (30 s association, 15 s dissociation, **Figure S7**). Solvent-corrected, double-referenced data were fit to a 1:1 binding model (below) using Biacore Insight Evaluation Software using the responses 5 s before the end of the injection phase.

$$P + L \rightleftharpoons PL \tag{7}$$
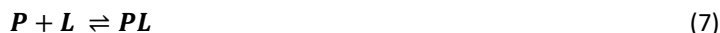
$$Response = \frac{R_{max} \cdot [L]}{[L] + K_d} + Offset \tag{8}$$

K_d	Equilibrium dissociation constant	(M)
R_{max}	Response at saturating ligand concentration	(RU)
$Offset$	Baseline offset	(RU)

	K_d	R_{max} (RU)	$Offset$ (RU)
Replicate 1	5.0E-04	44.7	-2.0
Replicate 2	6.1E-04	50.7	-2.0

Expression of human BCKDK. C-terminal hexa-histidine tagged human BCKDK (residues 34-412) was expressed in ExpiSf9 cells using the Bac-to-Bac Baculovirus system (Thermo Fisher). Cells were resuspended in 100 mM sodium phosphate buffer pH 7.5, 250 mM KCl, 500 mM NaCl, 200 mM arginine, 5 mM imidazole, 2 mM DTT supplemented with Dnase and complete EDTA-free protease inhibitors (Roche) and lysed by sonication. The lysate was cleared by centrifugation. Recombinant BCKDK was purified over Ni Sepharose 6 Fast Flow (GE Healthcare), followed by size exclusion chromatography using a Superdex 200 column (GE Healthcare). The peak fractions were pooled and concentrated to 5.5 mg/ml in 20 mM Hepes pH 7.5, 800 mM NaCl, 250 mM KCl, 300 mM arginine, and 1 mM TCEP, snap-frozen in liquid nitrogen, and stored at -80°C.

X-ray analysis of the BCKDK / compound 5 complex. Protein of human BCKDK (A34-I412)-GS-6×His at a concentration of 5.5 mg/mL in formulation 20 mM Hepes pH 7.5, 0.8 M NaCl, 250 mM KCl, 300 mM arginine, and 1 mM TCEP was set up in crystallization trials. Large enough crystals to use in soaking were obtained by sitting drop crystallization with a 2:1 ratio of protein solution and well solution containing 20-40 % EDO_P8K (Morpheus precipitant mix 2) and 0.1 M MB3 (Morpheus buffer system 3) 8.5 at 20 °C. A soaking drop (35 % EDO_PK8, 0.1 M MB3 pH 8.5), to which crystals were transferred, was prepared by adding a compound dissolved in DMSO to a final compound concentration of 6.5 mM. Crystals were soaked for 48 hrs and then flash-frozen in liquid nitrogen. Data

was collected at the European Synchrotron Radiation Facility, France, at beamline ID23-1. The crystal diffracted to a resolution of 2.64 Å but was somewhat anisotropic (**Table S7**). Data was processed using autoPROC with STARANISO and showed that the crystals belong to space group P4$_2$2$_1$2.[33, 34] The structure was solved by molecular replacement using the structure 1GKX as the search model.[35] An initial round of refinement was carried out in Refmac.[36] Several rounds of manual rebuilding in Coot and subsequent refinement in autoBUSTER were carried out prior to modeling of the ligand.[37, 38] Ligand restraints were generated with the program writedict.[39] Details from the data processing and refinement are reported in **Table S7**.

In vitro LC-MS assay for BCKDK. BCKDK (100 nM) was incubated with 20 µM or 400 µM of peptide substrate (RIGHHSTSDDSSAY; Innovagen) and 5 µM or 500 µM of ATP in assay buffer (20 nM HEPES, pH 7.5, 50 mM KCl, 5 mM MgCl2, 0.01% BSA, 1 mM glutathione) in a final volume of 10 µl. For testing compounds in concentration-response, 100 nl of compound dissolved in DMSO was included in the reactions (typically ten concentrations from 0.01 – 100 µM final assay concentration). Reactions were incubated for 24 hours at room temperature and stopped by the addition of 70 µl of 0.1% formic acid. For LC-MS analysis, samples (1 µl) were injected into an UPLC system (ACQUITY; Waters) containing an ACQUITY UPLC HSS T3 column (2.1 x 30 mm, 1.8 µm; Waters) kept at 40 °C at a flow rate of 1 ml/min in 0.2% acetonitrile with 0.1% formic acid. Analytes were separated using a linear gradient of 0.2-95% acetonitrile with 0.1% formic acid over 0.1-0.7 min and directed into a Xevo TQ-S triple quadrupole mass spectrometer (Waters) with an electrospray ionization source. The source temperature and desolvation temperature were set to 150 °C and 450 °C, respectively. Multiple reaction monitoring was used to measure the triply positively charged ions of the peptide substrate (511.6>253.1 m/z) and phosphopeptide product (538.2>253.1 m/z). Peak integration and quantification were done using MassLynx TargetLynx software (Waters). Curve fitting and calculation of pIC$_{50}$ were done using Genedata Screener (Genedata). Reported data are from three independent experiments.

Supplemental results

Identifying pockets in the X-ray structures used as input for the MD simulations. For p38-α (PDB id 5ETC, **Figure S1** and **Table S1**), we identified 16 pockets. The known allosteric pocket or lipid-binding pocket (LBP) was not detected among these pockets. The LBP is near the MAPK inset and involved in the kinase autophosphorylation and activation.[40] The absence of this pocket in the X-ray structure is due to a required structural shift of the αE-αF linker and α1L14-α2L14 (**Figure S3**).[41] However, Fpocket detected the D-groove and the non-canonical sites in p38-α. Although there has been no allosteric mechanism reported for any of these sites, they are functionally relevant for p38-α because they can bind specific proteins, such as TAB1, leading to kinase autophosphorylation and activation.[42]

In LFA-1, the allosteric pocket was also missing in the X-ray structure (PDB id 3F74, **Figure S1** and **Table S2**). This pocket would be visible only when helix 7 assumes an extended conformation, which varies by ~2.4 Å with respect to the bent conformation in 3F74 used here (**Figure S3**).[43] Allosteric inhibitors bind the extended helix 7 conformation, locking LFA-1 in a state with a low affinity towards the orthosteric binder (i.e., metal ion-dependent adhesion site, MIDAS).[43]

For GR, we identified eight pockets. However, the allosteric pocket in the AF-2 region of GR [44] is missing due to its shallow shape for binding co-regulator peptides (**Figure S1** and **Table S2**). Only in the cases of BCKDK and MAT2A, the known allosteric pockets are already present in the X-ray structures and identified by Fpocket (**Figure S1** and **Table S2**).

For BCKDK, we identified seven pockets (**Figure S1**). The allosteric pocket is located in the N-terminal domain, and small fragments targeting this site inhibit the kinase activity.[45] Although details of the allosteric mechanisms have not been fully elucidated, it has been proposed that the allosteric regulation also involves the lipoyl pocket.[45] Note that in the available BCKDK's X-ray structures, the lipoyl pocket has been resolved in a different conformation than in the case of the homologous protein pyruvate dehydrogenase kinase (PDHK). In BCKDK, the β-hairpin motifs occlude the lipoyl pocket, impeding its detection with Fpocket.

In MAT2A, Fpocket identified 29 pockets (**Figure S1**). The two allosteric pockets in MAT2A, one in each monomer, are nearby and often detected as a single large pocket. Targeting the allosteric pocket affects the active site, resulting in a decreased enzyme turnover.[46]

Additional figures
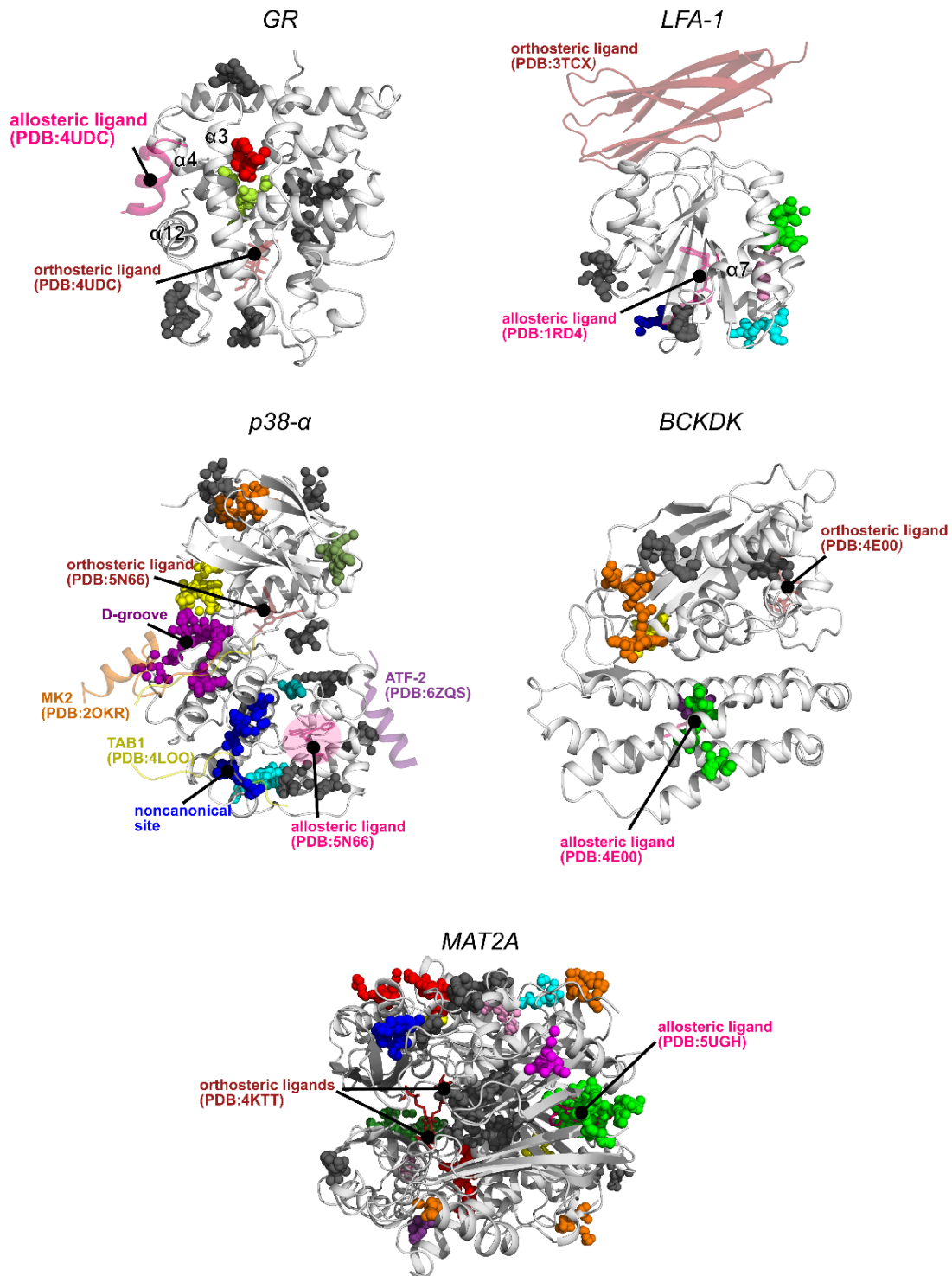
SI Figure 1. Pockets in *apo* X-ray structures. MAT2A, BCKDK, GR, LFA-1, and p38-α proteins are represented as a white cartoon representation, while pockets are represented as spheres. Grey pockets are only present in the X-ray structure and disappear in MD simulations, while colored pockets are also formed in MD simulations. Orthosteric and allosteric ligands and known protein partners are superimposed onto the apo structures and depicted as transparent sticks or cartoons.
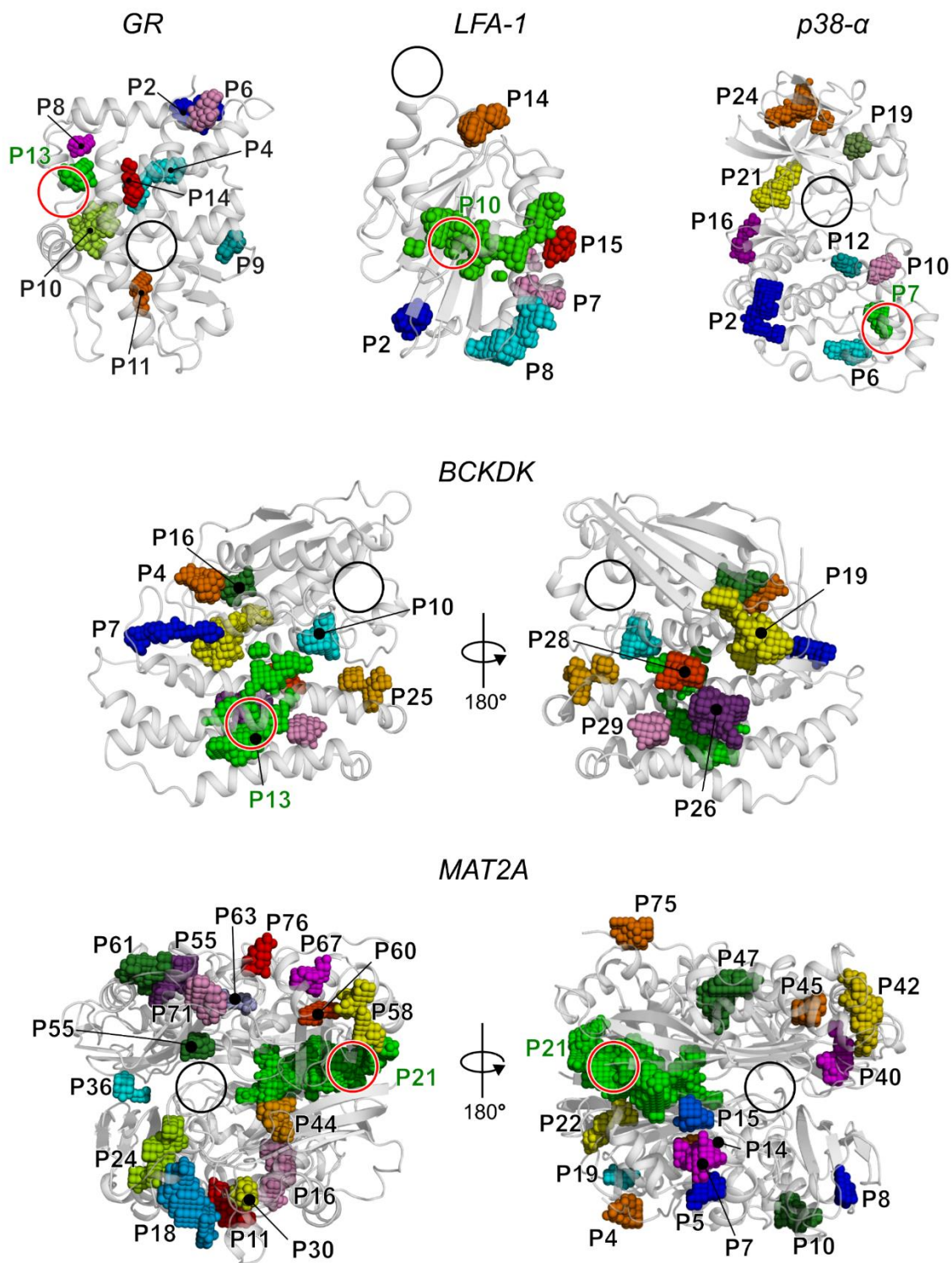
SI Figure 2. Pockets detected in MD simulations. GR, LFA-1, p38-α, BCKDK, and MAT2A proteins are represented as white cartoon, while pockets are represented as colored spheres. Only filtered pockets are shown (see Materials and Methods). The position of the orthosteric and allosteric sites are highlighted with black and red circles, respectively. Representative structures are extracted from MD simulations.
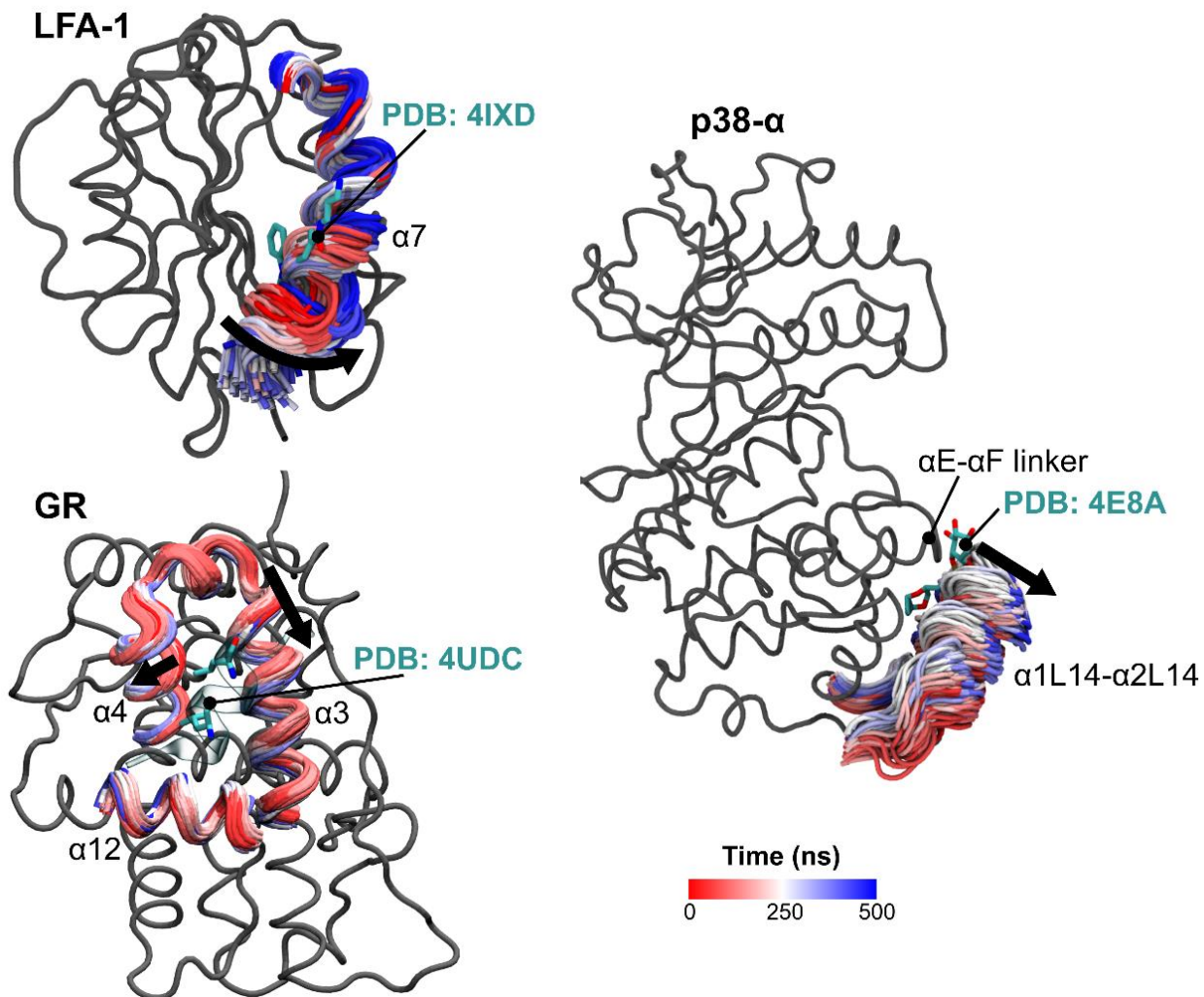
SI Figure 3. Representation of the time-dependent rearrangement of helices in LFA, GR, and p38-α. Early-stage frames are colored in red, while late stages are blue. For clarity, only helices relevant for opening allosteric pockets are represented. Black arrows show the main movements of the helices. Allosteric ligands are represented as cyan sticks and have been manually superimposed from the PDB ID reported in the figure directly on the MD trajectory.
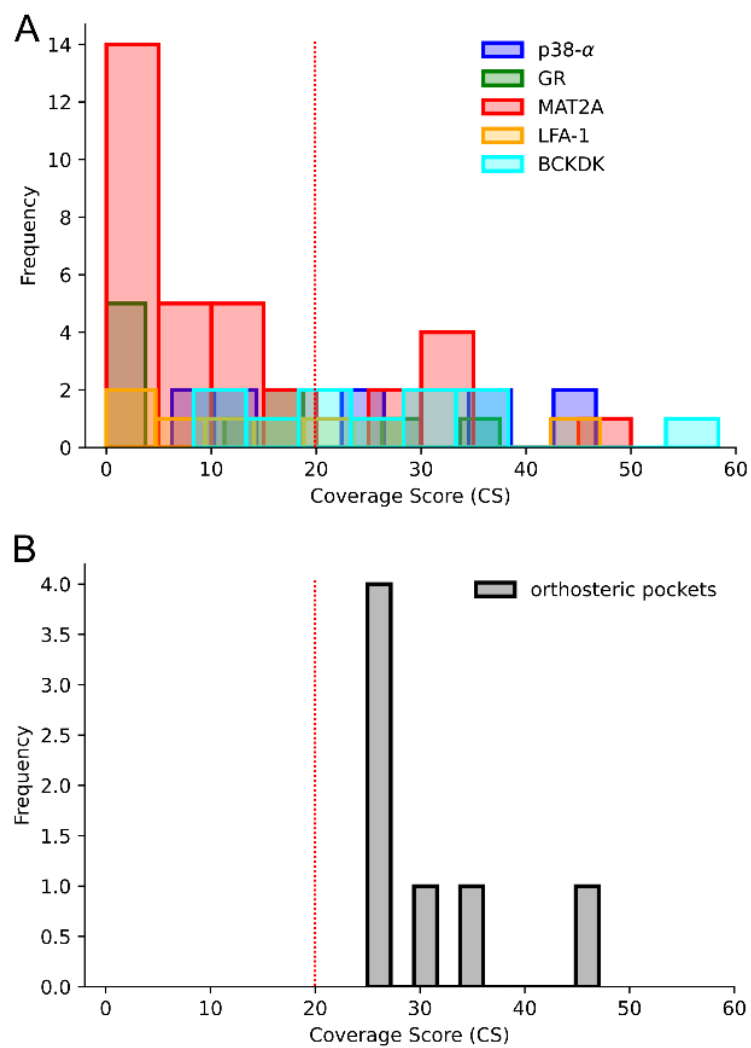
SI Figure 4. Distribution of the Coverage Score (CS). A) Distribution of the Coverage Score (CS) obtained via SCA of all the pockets found in the MD simulations. Each color corresponds to a system. B) Distribution of the Coverage Score (CS) for all the orthosteric pockets. The dotted red line represents the threshold used in this study (CS = 20%).
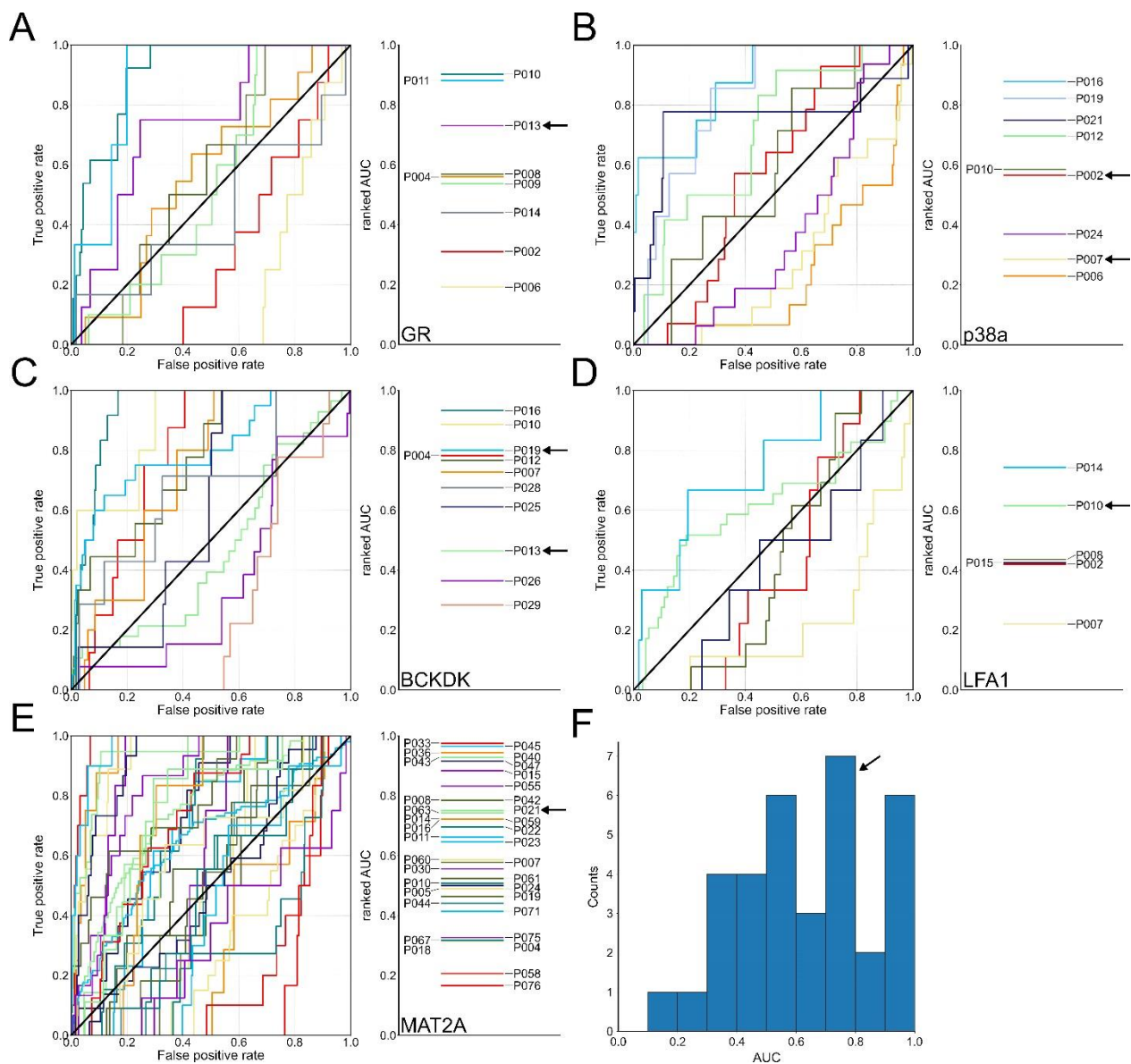
SI Figure 5. ROC analyses for the identified pockets using results from rigidity analysis. The ROC curves show the enrichment of residues having larger $\Delta G_{i,\mathrm{CNA}}$ values in the identified pockets (true positive rate) than other surface residues (false positive rate) for GR (A), p38-α (B), BCKDK (C), LFA-1 (D), and MAT2A (E). Right attached is the ranking based on calculated AUC values with arrows highlighting the allosteric/functional sites. F) In the case of MAT2A, the histogram shows the distribution of the AUC values.

14
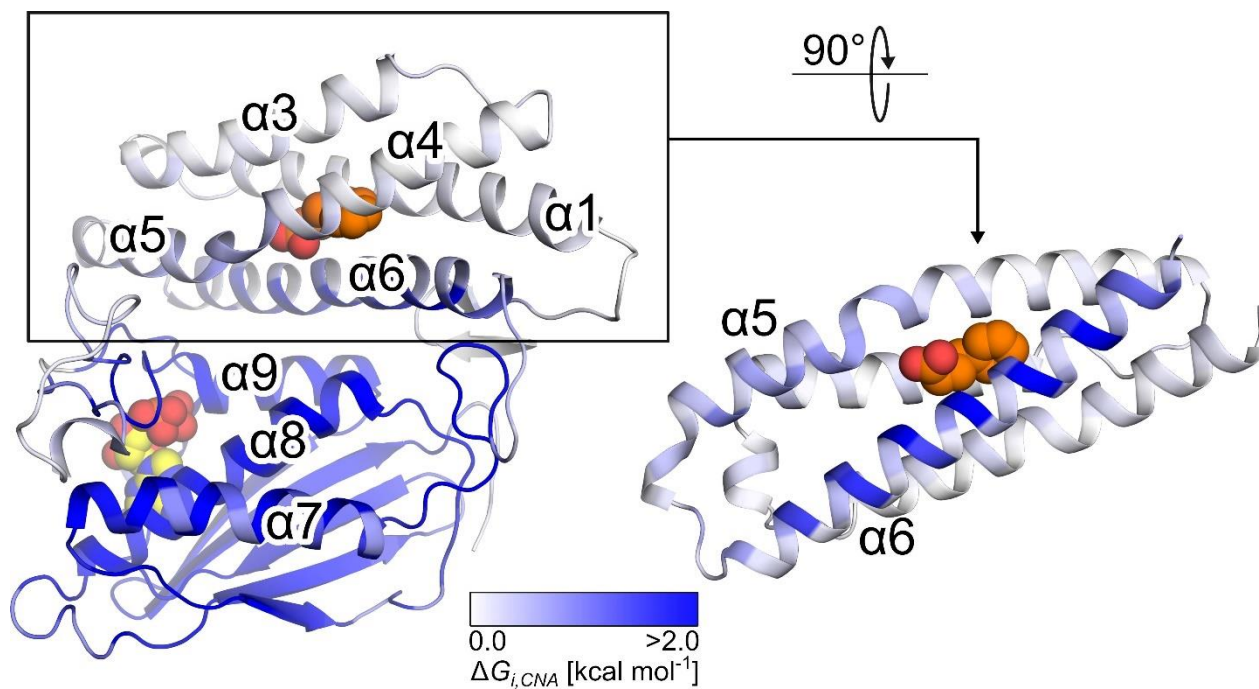
SI Figure 6. Altered structural stability in BCKDK from rigidity analysis. The per-residues free energies of altered structural stability $\Delta G_{i,\text{CNA}}$ are mapped on the structure of BCKDK (PDB id 1GKZ). Darker colors indicate larger $\Delta G_{i,\text{CNA}}$ values. Yellow spheres highlight the bound ADP in the orthosteric site. For depicting the position of the allosteric site, coordinates of the inhibitor were taken from the superimposition of the PDB id 3TZ2 and PDB id 1GKZ, with the inhibitor shown as orange spheres. The right site depiction highlights BCKDK's N-terminal domain with the two helices α5 and α6 that feel the perturbation when removing ADP from the structure.

SI Figure 7. *SPR sensorgrams and dose-response curve analyses for Compound 1 binding to MAT2A.* *Sensorgrams are colored according to concentration (667 μM top concentration, 1.5-fold serial dilution). The binding affinity of Compound 1 was estimated as 550 ± 75 μM using these data.*
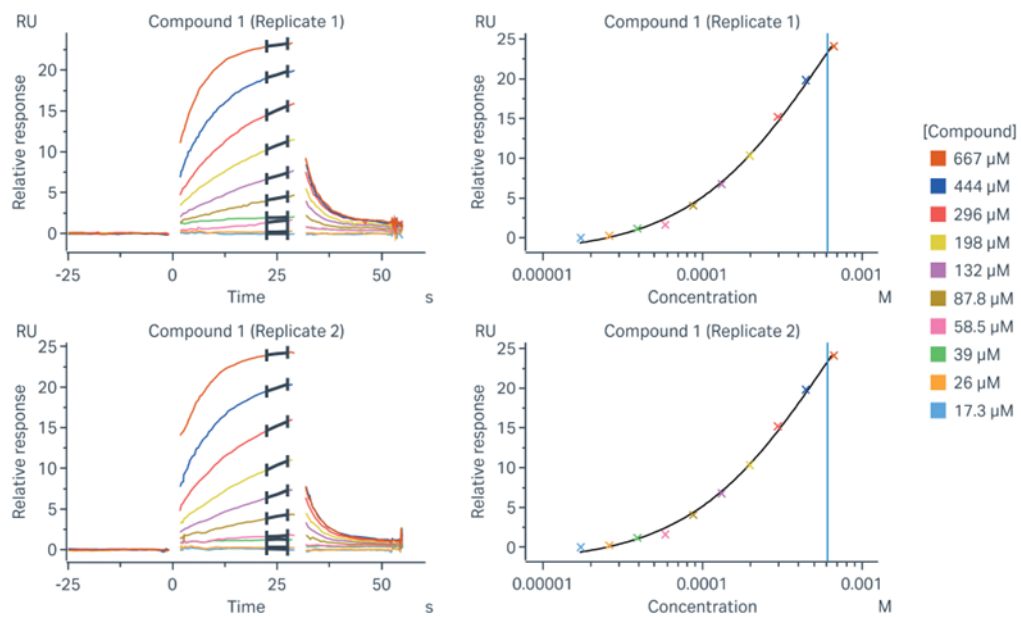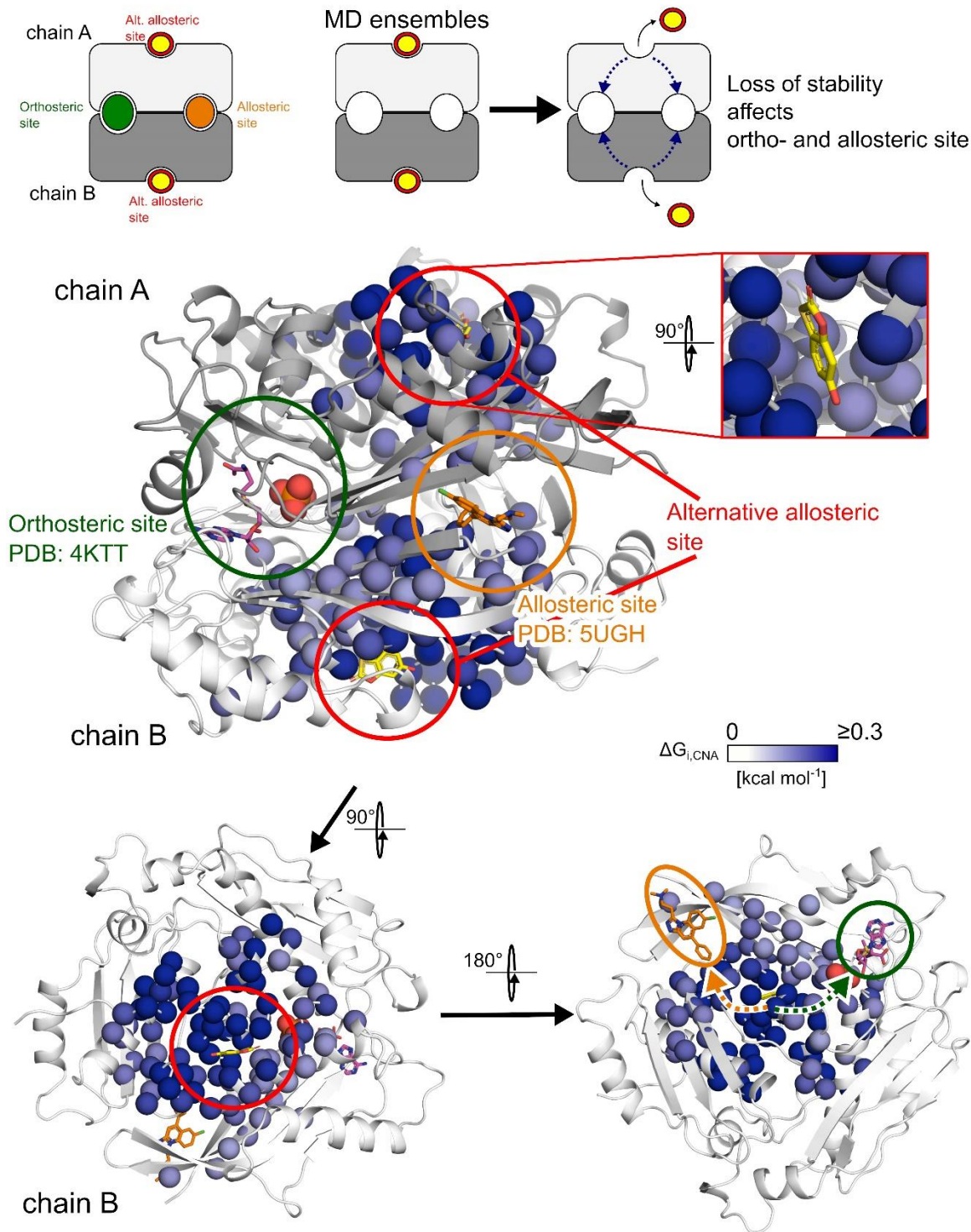
SI Figure 8. Stability changes in MAT2A deduced from the perturbation approach. Top: MD simulations were performed starting from the new X-ray structure of the MAT2A dimer in complex with Compound 1 (PDB id 8OOG, see upper scheme). The removal of compound 1 impacts the structural stability affecting both the orthosteric and the allosteric pockets. The figures in the middle and the bottom show the per-residues free energies of altered structural stability $\Delta G_{i,CNA}$ mapped on the structure of MAT2A (PDB id 8OOG). Darker colors indicate larger $\Delta G_{i,CNA}$ values. Orthosteric (PDB id 4KTT) and allosteric (PDB id 5UGH) site ligands are shown as magenta and orange sticks, respectively and are superimposed on the 8OOG structure.

A
PDB: 1GKZ
α7 α8
Met333 ADP
α9

Superimpose Cpd5 →

PDB: 1GKZ
Met333 ADP
Compound 5
(Superimposed)

B
PDB: 1GKX
P0X-ray

Superimpose Cpd5 →

PDB: 1GKX
P0X-ray
Compound 5
(Superimposed)

C
MD frame
with loop
Met333
P10

Superimpose Cpd5 →

MD frame
with loop
Met333
Compound 5
(Superimposed)
P10

D
MD frame
without loop
P0MD

Superimpose Cpd5 →

MD frame
without loop
P0MD
Compound 5
(Superimposed)

SI Figure 9. Representation of the new allosteric pocket identified in our study in the BCKDK system. Four different structures are compared to show how the loop region (residues 297-340, green cartoon) affects the pocket formation in between α–helices 7-8 (namely P0$_{X-ray}$ in X-ray structures or P0$_{MD}$ in MD snapshots). A) In the X-ray structure PDB id 1GKZ, only the orthosteric ATP pocket is formed (pink surface). In this X-ray structure, ADP is bound to the ATP pocket, and Met333 (green sticks) binds in between α–helices 7-8, preventing the formation of P0$_{X-ray}$. On the right, compound 5 is superimposed on the X-ray structure 1GKZ. B) In the X-ray structure 1GKX, both the orthosteric ATP pocket and the P0$_{X-ray}$ are formed (pink and cyan surfaces, respectively). P0$_{X-ray}$ is formed because the loop corresponding to Met333 is not solved. On the right, compound 5 is superimposed on the X-ray structure 1GKX and nicely overlaps with P0$_{X-ray}$. C) In the MD simulations performed on the full-length structure (loop is entirely modeled, green cartoon), only the pocket P10 is formed (lemon surface). The P0$_{MD}$ is not formed because Met333 occupies that site, as for the X-ray 1GKZ. On the right, compound 5 is superimposed on the MD frame, clashing with Met333. D) Pocket detection is performed on MD simulations after manual removal of the loop (residues 297-340). This time the P0$_{MD}$ is formed (cyan surface). On the right, compound 5 is superimposed on the MD frame.
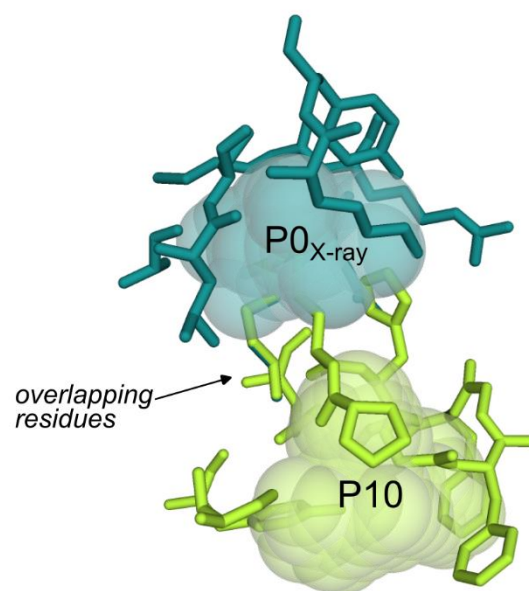
SI Figure 10. Overlapping residues between P0$_{X-ray}$ and P10. The pocket P0$_{X-ray}$ in the PDB structure 1GKZ is represented as cyan spheres, while the residues surrounding P0$_{X-ray}$ are depicted as cyan sticks. P10 is represented as lemon spheres, while residues surrounding P10 are depicted as lemon sticks.
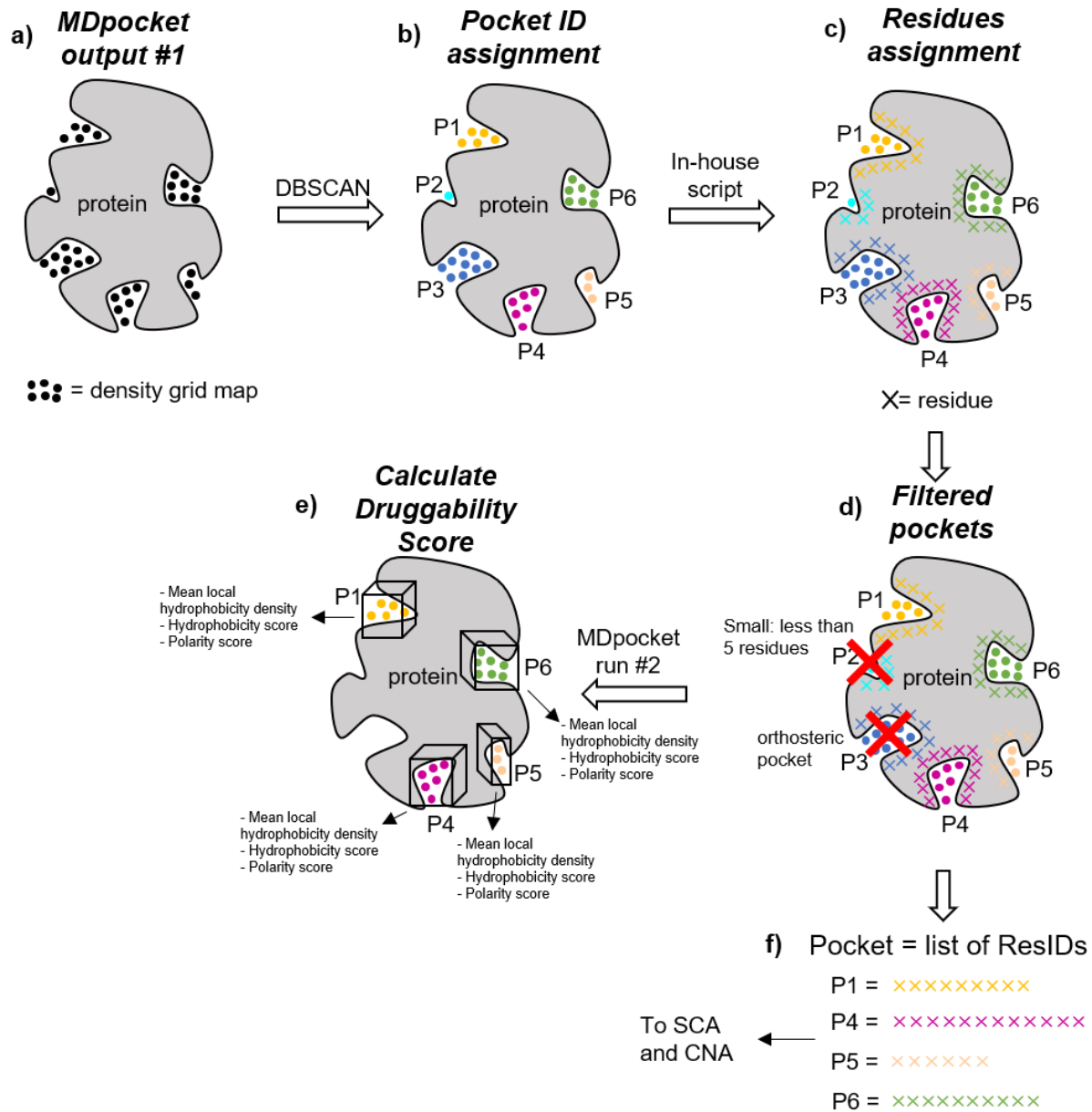
SI Figure 11. Workflow of the pocket detection protocol. A) The input of the first MDpocket run consists of a density map. We considered the map at Isovalues = 2 for all the systems except for LFA-1, where we used Isovalues = 1. B) The DBSCAN algorithm is applied to cluster the data points from the density map (see step A) into separate pockets. Each pocket is assigned an ID. C) An in-house script is employed to extract the protein's residues surrounding each pocket. D) The known orthosteric pocket and small pockets (surrounded by less than five residues) were removed. E) A second MDpocket run is carried out to retrieve the mean local hydrophobic density, hydrophobicity score, and polarity score for each pocket. For this run, we generated a PDB file (one PDB for each pocket) containing the dummy atoms at the grid points' position that overlap with the grid point in the clustered density map (see step B). The final mean local hydrophobic density, the hydrophobicity score, and the polarity score values of each pocket are used to estimate the druggability score from averaging over the MD trajectory. F) Each filtered pocket is translated into a list of residues IDs (resIDs) assigned in step C. The list of residues is used for the SCA and CNA analysis.

BCKDK

p38-α

LFA-1

MAT2A

GR
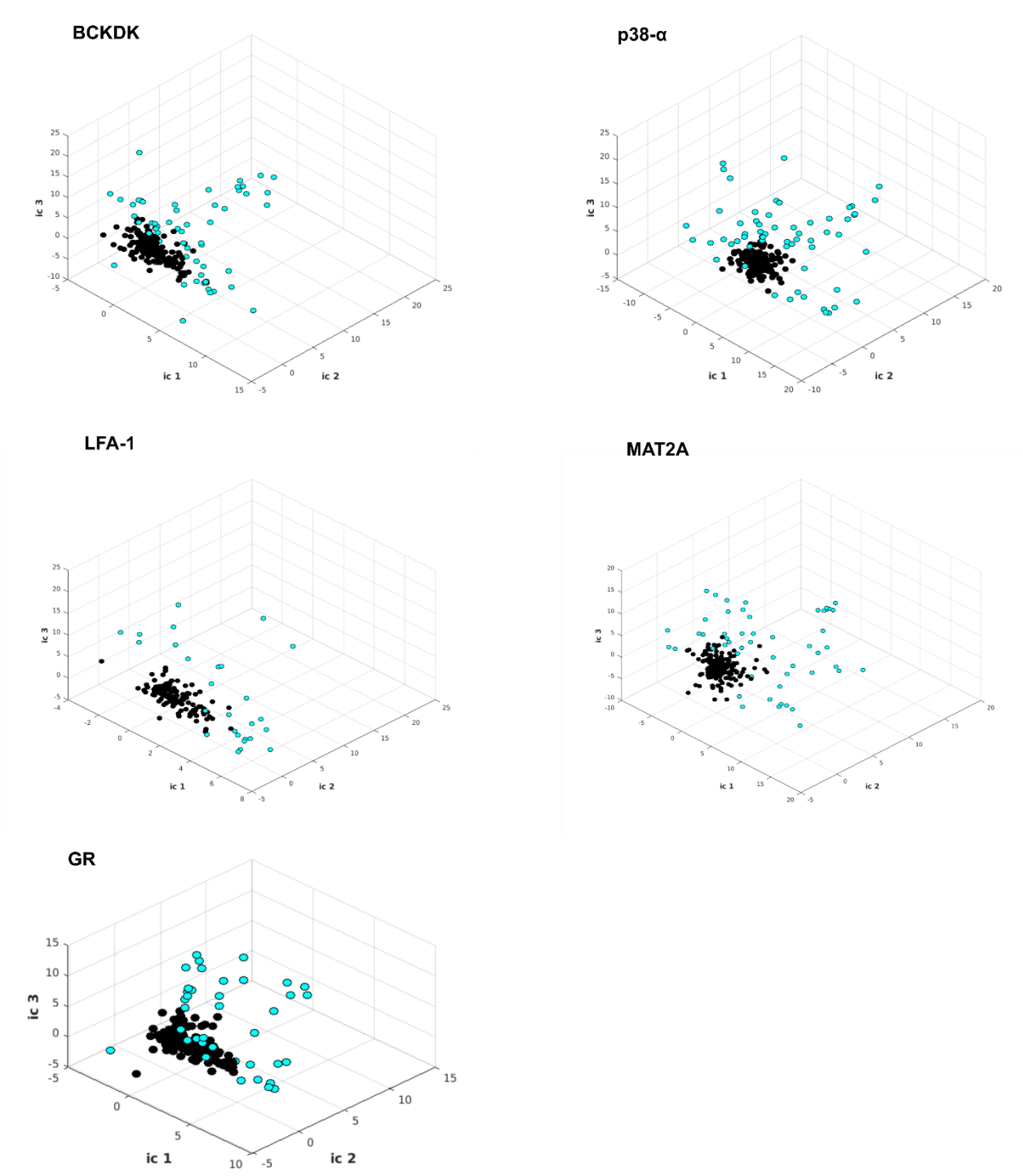
SI Figure 12. Independent Component Analysis (ICA) decomposition. The weakly correlated residues are grouped close to the origins of the axes and are depicted as black points. The cyan points correspond to evolutionarily correlated positions.
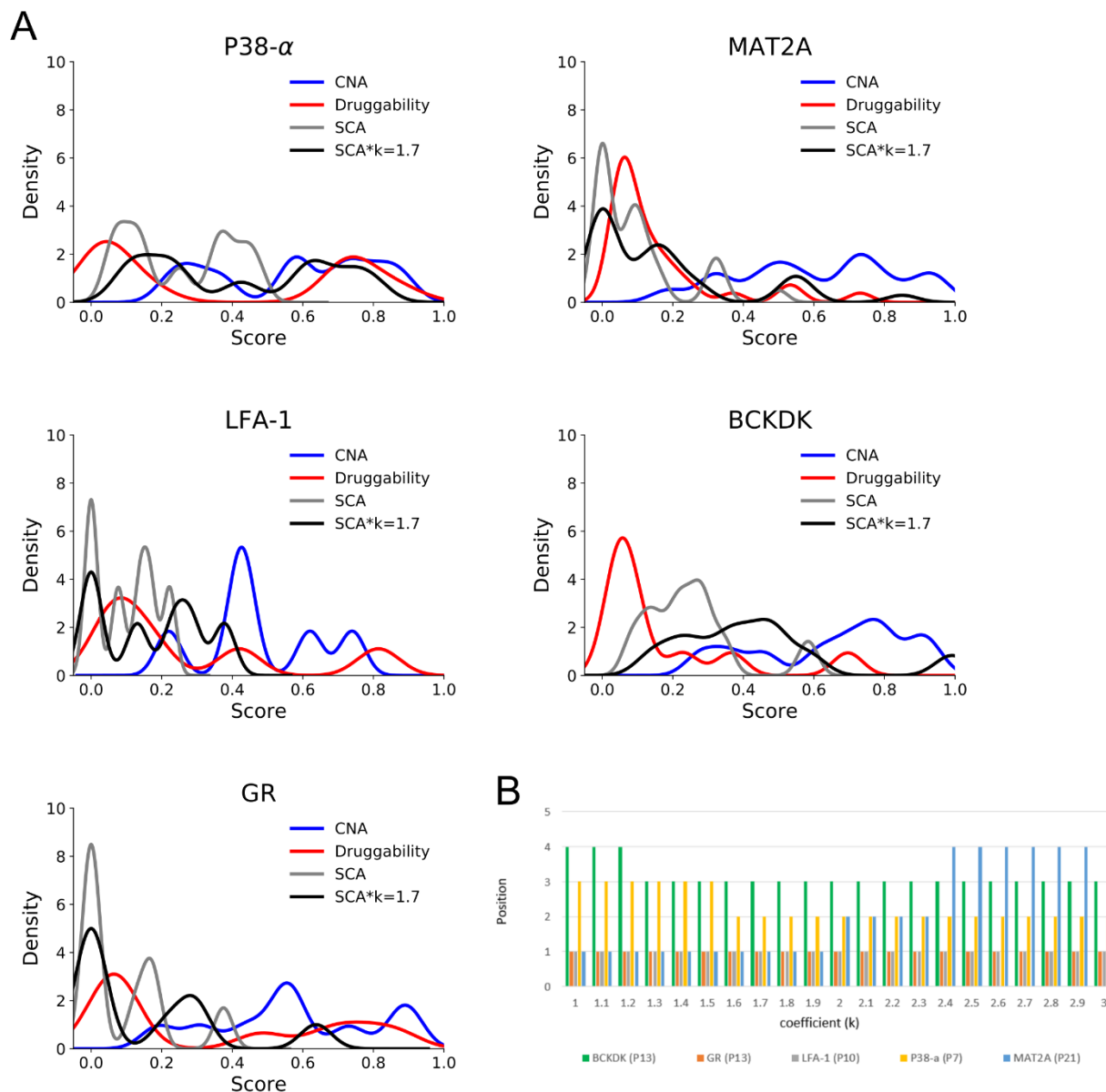
SI Figure 13. Effects of k coefficient for SCA. A) Distribution of the single scores for each system. The blue, red, grey, and black lines correspond to the CNA score, Druggability score, SCA score and SCA score multiplied by 1.7, respectively. B) The bar plot shows the final ranking of the allosteric pockets using coefficients *k* from 1 to 3 in steps of 0.1. The green, orange, grey, yellow, and blue bars correspond to the ranking of allosteric pockets in BCKDK, GR, LFA-1, p38-α, and MAT2A, respectively.

Additional tables

Table S1. Number of pockets identified in X-ray and MD simulations.

Protein	X-ray APO [PDB id] [a]	MD simulations total [a]	MD simulations filtered [a]	MD simulations new pockets [a]
LFA-1	7 [3F74 chain C]	15	6	3 (P10,P14,P15)
p38-α	16 [5ETC]	25	9	2 (P7,P10)
BCKDK	7 [1GKX]	28	10	5 (P10,P16,P25,P28,P29)
GR	8 [4UDC]	15	9	7 (P2,P4,P6,P8,P9,P11,P13)
MAT2A	29 [6FAJ]	74	31	15 (P5,P14,P15,P18,P22, P24, P36,P40, P42,P44,P47, P60,P61,P63,P67)

[a] Orthosteric pocket is removed.

Table S2. Volumes of the allosteric pockets.

Protein	Volume HOLO X-ray [PDB id] [a]	Volume APO X-ray [PDB id] [a]	Volume MD simulations [b, c]
LFA-1	972 Å^3 [4IXD]	ND [3F74 chain C]	243 ± 129 Å^3
p38-α	1077 Å^3 [3E8A]	ND [5ETC]	218 ± 108 Å^3
BCKDK	373 Å^3 [3TZ0]	370 Å^3 [1GKX]	457 ± 170 Å^3
GR	ND	ND [4UDC]	181 ± 65 Å^3
MAT2A	1612 Å^3 [5UGH]	1694 Å^3 [6FAJ]	2285 ± 383 Å^3

[a] Calculated by Fpocket using default settings.
[b] Calculated by MDpocket using default settings.
[c] Average ± standard deviation.

Table S3. Residue conservation and co-evolution values were calculated for each pocket. Residue conservation is calculated according to the Kullback-Leibler relative entropy, and in this table, we report the pocket's average and median (3rd and 4th column, respectively). In the 5th column, the maximum entropy value within the pocket is reported and in the 6th the coverage score from SCA.

GR

PocketID	Annotation	Entropy Average	Entropy Median	Entropy Max	Coverage Score (from SCA)
2		0.77	0.74	1.14	0
4		0.69	0.71	1.23	0
6		0.96	0.85	2.03	12.5
8		0.77	0.81	1.15	0
9		0.89	0.77	2.17	0
10		1.19	0.96	4.02	17.65
11		0.81	0.76	1.20	0
12	Orthosteric	1.38	1.28	4.21	29.55
13	Allosteric	2.05	2.13	3.16	37.5
14		1.25	0.71	4.21	16.67

LFA-1

PocketID	Annotation	Entropy Average	Entropy Median	Entropy Max	Coverage Score (from SCA)
2		1.00	0.58	2.54	0
3	Orthosteric	1.25	1.05	2.38	47.06
7		1.11	0.99	2.48	22.22
8		0.93	0.84	1.59	7.69
10	Allosteric	0.95	0.75	2.67	13.79
11	Orthosteric	1.28	1.28	2.74	25
14		0.58	0.55	1.03	0
15		0.83	0.76	1.40	16.67

MAT2A

PocketID	Annotation	Entropy Average	Entropy Median	Entropy Max	Coverage Score (from SCA)
2	Orthosteric	1.30	1.36	2.78	26.76
4		0.65	0.53	1.05	0
5		1.20	1.25	1.67	0
7		1.26	1.25	2.41	11.11
8		0.37	0.29	0.64	0
10		0.52	0.47	0.89	0
11		0.79	0.87	1.28	7.69
14		1.47	1.27	2.78	9.09
15		1.27	1.09	2.41	16.67
16		1.25	1.12	2.90	31.25
18		0.74	0.43	1.94	10
19		0.86	0.84	1.29	50
21	Allosteric	1.20	0.99	3.30	11.29
22		0.98	0.77	2.39	15.38
23	Orthosteric	1.48	1.50	2.90	26.42

24	0.82	0.52	2.47	4.55
30	0.81	0.76	1.28	0
36	0.56	0.42	1.65	0
40	1.08	0.75	2.66	13.33
42	0.83	0.51	2.18	7.14
44	1.34	1.29	2.05	33.33
45	1.04	0.89	2.23	0
47	1.42	1.20	3.30	31.25
55	0.66	0.55	1.28	6.67
58	1.26	1.26	2.41	10
59	0.67	0.38	2.30	0
60	1.36	1.27	1.73	9.09
61	0.48	0.33	0.89	0
63	1.64	1.58	2.73	33.33
67	1.15	1.25	1.67	0
71	0.44	0.40	0.90	0
75	0.60	0.50	1.05	0
76	0.75	0.69	1.59	0

P38-α

PocketID	Annotation	Entropy Average	Entropy Median	Entropy Max	Coverage Score (from SCA)
2	noncanonical	1.26	1.20	2.53	35.71
6		1.21	0.81	2.51	46.67
7	Allosteric	1.50	1.15	3.17	37.5
10		1.91	2.24	2.97	42.86
12		2.68	2.75	3.96	25
14	ATP site	1.80	1.59	3.23	26.19
16	D-groove	1.14	1.02	2.23	12.5
19		1.32	1.11	2.34	14.29
21		0.89	0.82	1.93	7.69
24		0.64	0.29	1.71	6.25

BCKDK

PocketID	Annotation	Entropy Average	Entropy Median	Entropy Max	Coverage Score (from SCA)
4		0.99	0.98	1.78	12.5
7		0.88	0.73	1.91	20
10	New Pocket	1.62	1.65	2.63	58.33
12	ATP site	1.43	1.27	2.92	35.29
13	Allosteric	1.30	1.07	2.59	34.48
16		1.28	1.25	2.24	8.33
19		1.09	0.99	2.72	28.57
25		0.98	0.71	2.49	25
26		1.14	1.14	2.39	15.38
28		1.23	1.04	2.39	28.57
29		1.02	1.00	1.82	22.22

Table S4. Pearson correlation (R^2 values) of per-residue free energies computed by CNA across all ten independent MD ensembles.

GR[1,2]

MD run	1	2	3	4	5	6	7	8	9	10
1		0.97	0.98	0.97	0.96	0.98	0.95	0.98	0.97	0.97
2	0.97		0.98	0.98	**0.94**	0.97	**0.94**	0.98	0.95	**0.94**
3	0.98	0.98		0.97	0.96	0.98	0.96	0.99	0.95	0.95
4	0.97	0.98	0.97		0.96	0.98	0.95	0.97	0.95	0.96
5	0.96	**0.94**	0.96	0.96		0.97	0.97	0.95	0.95	0.96
6	0.98	0.97	0.98	0.98	0.97		0.97	0.98	0.96	0.96
7	0.95	**0.94**	0.96	0.95	0.97	0.97		0.95	0.96	0.95
8	0.98	0.98	0.99	0.97	0.95	0.98	0.95		**0.94**	**0.94**
9	0.97	0.95	0.95	0.95	0.95	0.96	0.96	**0.94**		0.97
10	0.97	**0.94**	0.95	0.96	0.96	0.96	0.95	**0.94**	0.97	

LFA-1[1,2]

MD run	1	2	3	4	5	6	7	8	9	10
1		0.95	0.80	0.96	0.92	0.98	0.91	0.98	0.99	0.93
2	0.95		0.87	0.99	0.96	0.99	0.98	0.97	0.93	0.99
3	0.80	0.87		0.86	0.90	0.84	0.89	0.86	**0.75**	0.85
4	0.96	0.99	0.86		0.94	0.99	0.96	0.97	0.94	0.98
5	0.92	0.96	0.90	0.94		0.95	0.98	0.96	0.87	0.96
6	0.98	0.99	0.84	0.99	0.95		0.96	0.99	0.97	0.97
7	0.91	0.98	0.89	0.96	0.98	0.96		0.95	0.87	0.99
8	0.98	0.97	0.86	0.97	0.96	0.99	0.95		0.97	0.95
9	0.99	0.93	**0.75**	0.94	0.87	0.97	0.87	0.97		0.91
10	0.93	0.99	0.85	0.98	0.96	0.97	0.99	0.95	0.91	

MAT2A[1,2]

MD run	1	2	3	4	5	6	7	8	9	10
1		**0.81**	0.87	0.94	0.90	0.83	0.92	0.95	0.87	0.89
2	**0.81**		0.91	0.86	0.93	0.91	0.87	0.92	0.94	0.88
3	0.87	0.91		0.94	0.94	0.93	0.95	0.92	0.96	0.90
4	0.94	0.86	0.94		0.95	0.92	0.98	0.95	0.93	0.93
5	0.90	0.93	0.94	0.95		0.95	0.94	0.96	0.95	0.96
6	0.83	0.91	0.93	0.92	0.95		0.91	0.90	0.94	0.92
7	0.92	0.87	0.95	0.98	0.94	0.91		0.94	0.96	0.91
8	0.95	0.92	0.92	0.95	0.96	0.90	0.94		0.93	0.95
9	0.87	0.94	0.96	0.93	0.95	0.94	0.96	0.93		0.90
10	0.89	0.88	0.90	0.93	0.96	0.92	0.91	0.95	0.90	

p38-$\alpha^{1,2}$

MD run	1	2	3	4	5	6	7	8	9	10
1	⬛	0.98	0.94	0.94	0.97	0.97	0.91	0.95	0.90	0.93
2	0.98	⬛	0.93	0.93	0.98	0.95	0.92	0.93	0.90	0.92
3	0.94	0.93	⬛	0.89	0.95	0.92	0.86	0.94	0.82	0.96
4	0.94	0.93	0.89	⬛	0.90	0.94	0.93	0.86	0.91	0.90
5	0.97	0.98	0.95	0.90	⬛	0.94	0.88	0.94	0.85	0.94
6	0.97	0.95	0.92	0.94	0.94	⬛	0.90	0.94	0.91	0.91
7	0.91	0.92	0.86	0.93	0.88	0.90	⬛	0.82	0.94	0.84
8	0.95	0.93	0.94	0.86	0.94	0.94	0.82	⬛	0.80	0.92
9	0.90	0.90	0.82	0.91	0.85	0.91	0.94	0.80	⬛	**0.78**
10	0.93	0.92	0.96	0.90	0.94	0.91	0.84	0.92	**0.78**	⬛

BCKDK^{1,2}

MD run	1	2	3	4[3]	5	6	7	8	9	10[3]
1	⬛	0.92	0.90	0.57	0.94	0.98	0.94	0.94	0.94	0.68
2	0.92	⬛	0.86	0.55	0.94	0.91	0.87	0.91	0.94	0.79
3	0.90	0.86	⬛	0.53	**0.82**	0.89	0.93	0.91	0.89	0.64
4[3]	0.57	0.55	0.53	⬛	0.51	0.54	0.59	0.42	0.52	0.27
5	0.94	0.94	**0.82**	0.51	⬛	0.94	0.85	0.92	0.92	0.79
6	0.98	0.91	0.89	0.54	0.94	⬛	0.95	0.94	0.93	0.67
7	0.94	0.87	0.93	0.59	0.85	0.95	⬛	0.91	0.91	0.60
8	0.94	0.91	0.91	0.42	0.92	0.94	0.91	⬛	0.92	0.73
9	0.94	0.94	0.89	0.52	0.92	0.93	0.91	0.92	⬛	0.77
10[3]	0.68	0.79	0.64	0.27	0.79	0.67	0.60	0.73	0.77	⬛

[1] p-value < 0.001.

[2] Lowest R^2 value is highlighted by gray background color.

[3] Trajectory was rejected from further analysis as the orthosteric site ligand did not bind stably during the simulation.

Table S5. Number of surface-exposed residues

System	Total number of residues	Number of surface residues[1,2,3]
MAT2A	760	448 (59%)
BCKDK	341	247 (72%)
LFA-1	179	108 (60%)
GR	248	169 (68%)
P38-α	352	247 (70%)

[1] A residue is surface exposed if the SASA is ≥ 3.5 Å.

[2] Orthosteric site residues are ignored because perturbation takes place at these sites.

[3] Percentages in parentheses are the proportion of surface residues to the total number of residues.

Table S6. Comparison of Scoring Combination. Values indicate the pocket's ranking estimated according to the one-parameter model (DS, rigidity analysis, and SCA), two-parameter model (rigidity analysis-SCA, SCA-DS, and rigidity analysis-DS), and three-parameter model (rigidity analysis-SCA-DS). Light green background indicates positions within the first third of the ranking ($R_{33\%}$). Dark green background highlights the best results.

	DS	Rigidity analysis	SCA	Rigidity analysis-SCA	SCA-DS	Rigidity analysis-DS	Rigidity analysis -SCA-DS
GR Allosteric	1	3	1	1	1	2	1
LFA-1 Allosteric	1	2	3	1	1	1	1
BCKDK Allosteric	2	8	2	6	3	5	3
MAT2A Allosteric	1	9	9	11	2	1	1
p38-α Allosteric	4	8	3	7	2	3	2
p38-α noncanonical	5	6	4	2	5	8	6
p38-α D-groove	6	1	7	4	9	5	8
BCKDK New Pocket (P0$_{MD}$)	7	2	1	1	2	3	2
MAT2A New Pocket (P47)	9	4	5	1	5	5	2
# Positive in $R_{33\%}$*	4	4	4	3	5	4	5
Average Rank*	2.85	5.28	4.14	4.57	3.28	3.57	3.14
# Positive in $R_{33\%}$**	4	5	5	4	6	5	6
Average Rank**	3.37	4.87	3.75	4.12	3.12	3.5	3

* Values calculated without the BCKDK's new pocket (P0$_{MD}$) and MATA2's new pocket (P47).

** Values calculated with the BCKDK's new pocket (P0$_{MD}$) and without MATA2's new pocket (P47).

Table S7. Crystallographic data collection and refinement statistics.

	BCKDK: Compound 5	MAT2A: Compound 1
PDB accession code	7ZPE	8OOG
Synchrotron source	ID23-1/ESRF	PX1/Soleil Synchrotron
Space group:	$P4_22_12$	I222
Cell constants: a, b, c (Å)	129.6, 129.6, 73.2	67.9, 94.2, 117.0
Resolution range (Å)	91.64 – 2.64	73.37 – 1.38
Highest resolution shell (Å)	2.76 – 2.64	1.41 – 1.38
Ellipsoidal resolution limit (direction) (Å)	2.63 (a*), 2.63 (b*), 2.80 (c*)	
Completeness ellipsiodal (%)	95.5 (57.2)	
Completeness overall (%)	90.4 (35.7)	98.7 (97.2)
Total number of observations	196738	466961
Reflections, unique	17091	75517
Redundancy	11.5 (12.6)	6.2 (6.4)
R merge $_{overall}$ [1]	0.068 (2.45)	0.070 (1.097)
I/SigI[2]	19.8 (1.3)	13.3 (2.1)
CC1/2	0.99 (0.885)	0.999 (0.685)
R value $_{overall}$ (%)[3]	21.8	19.0
R value $_{free}$ (%)[4]	24.9	20.0
Bond lengths (Å)	0.009	0.008
Bond angles (º)	0.094	0.997
In most favored regions (%)[5]	92.1	90.3
In additional allowed regions (%)[5]	7.5	9.4
In generously regions (%)[5]	0.4	0.3
In disallowed regions (%)[5]	0.0	0.0

[1] $R_{merge} = \Sigma_{hkl}\left[\left(\Sigma_i |I_i - \langle I\rangle|\right)/\Sigma_i I_i\right]$
[2] I/sigI avg is the mean I/sig for the unique reflections in the output file
[3] $R_{value} = \Sigma_{hkl} ||F_{obs}| - |F_{calc}|| / \Sigma_{hkl} |F_{obs}|$
[4] R_{free} is the cross-validation R factor computed for the test set of 5 % of unique reflections
[5] Ramachandran statistics as defined by PROCHECK[47]

Table S8. Potency of compound 5 against BCKDK measured in the in vitro LC-MS assay

Condition	[peptide] (μM)	[ATP] (μM)	pIC50 (mean ± SD, n=3)
Low peptide, Low ATP	20	5	4.90 ± 0.23
Low peptide, High ATP	20	500	5.07 ± 0.10
High peptide, High ATP	400	500	5.01 ± 0.17

References

1   W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926–935.
2   M. Parrinello and A. Rahman, *J. Appl. Phys.*, 1981, **52**, 7182–7190.
3   G. Bussi, D. Donadio and M. Parrinello, *J. Chem. Phys.*, 2007, **126**, 14101.
4   B. Hess, H. Bekker, H. J. C. Berendsen and J. G. E. M. Fraaije, *J. Comput. Chem.*, 1997, **18**, 1463–1472.
5   T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089–10092.
6   M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindahl, *SoftwareX*, 2015, **1–2**, 19–25.
7   V. Le Guilloux, P. Schmidtke and P. Tuffery, *BMC Bioinformatics*, 2009, **10**, 168.
8   P. Schmidtke, A. Bidon-chanal, F. J. Luque and X. Barril, *Bioinformatics*, 2011, **27**, 3276–3285.
9   P. Schmidtke and X. Barril, *J. Med. Chem*, 2010, **53**, 5858–5867.
10   O. Rivoire, K. A. Reynolds and R. Ranganathan, *PLoS Comput. Biol.*, 2016, **12**, 1–26.
11   K. A. Reynolds, R. N. Mclaughlin and R. Ranganathan, *Cell*, 2011, **147**, 1564–1575.
12   M. Novinec, M. Korenč, A. Caflisch, R. Ranganathan, B. Lenarčič and A. Baici, *Nat. Commun.*, 2014, **5**, 3287.
13   T. Cover and J. Thomas, *John Wiley Sons*.
14   A. Panjkovich and X. Daura, *BMC Struct. Biol.*, 2010, **10**, 9.
15   N. Halabi, O. Rivoire, S. Leibler and R. Ranganathan, *Cell*, 2009, **138**, 774–786.
16   M. K. Hjortness, L. Riccardi, A. Hongdusit, P. H. Zwart, B. Sankaran, M. De Vivo and J. M. Fox, *Biochemistry*, 2018, **57**, 6443–6451.
17   S. R. Eddy, *PLOS Comput. Biol.*, 2011, **7**, e1002195.
18   C. Pfleger, P. C. Rathi, D. L. Klein, S. Radestock and H. Gohlke, *J. Chem. Inf. Model.*, 2013, **53**, 1007–1015.
19   D. J. Jacobs, A. J. Rader, L. A. Kuhn and M. F. Thorpe, *Proteins Struct. Funct. Bioinforma.*, 2001, **44**, 150–165.
20   B. I. Dahiyat, D. Benjamin Gordon and S. L. Mayo, *Protein Sci.*, 1997, **6**, 1333–1337.
21   A. J. Rader, B. M. Hespenheide, L. A. Kuhn and M. F. Thorpe, *Proc. Natl. Acad. Sci.*, 2002, **99**, 3540–3545.
22   S. Radestock and H. Gohlke, *Eng. Life Sci.*, 2008, **8**, 507–522.
23   A. J. Rader, *Phys. Biol.*, 2009, **7**, 016002.
24   S. Radestock and H. Gohlke, *Proteins Struct. Funct. Bioinforma.*, 2011, **79**, 1089–1108.
25   B. M. Hespenheide, A. J. Rader, M. F. Thorpe and L. A. Kuhn, *J. Mol. Graph. Model.*, 2002, **21**, 195–207.
26   C. Pfleger, A. Minges, M. Boehm, C. L. McClendon, R. Torella and H. Gohlke, *J. Chem. Theory Comput.*, 2017, **13**, 6343–6357.
27   C. Pfleger, S. Radestock, E. Schmidt and H. Gohlke, *J. Comput. Chem.*, 2013, **34**, 220–33.
28   M. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. W. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin and K. S. Wilson, *Acta Crystallogr. D. Biol. Crystallogr.*, 2011, **67**, 235–42.
29   J. Navaza, *Acta Crystallogr. Sect. D Biol. Crystallogr.*, 2001, **57**, 1367–1372.
30   G. Bricogne, E. Blanc, M. Brandl, C. Flensburg, P. Keller, W. Paciorek, P. Roversi, A. Sharff, O. S. Smart, C. Vonrhein and T. O. Womack. BUSTER v.2.11.6., Global Phasing Ltd., Cambridge, UK.
31   P. Emsley, B. Lohkamp, W. G. Scott and K. Cowtan, *Acta Crystallogr. Sect. D Biol. Crystallogr.*, 2010, **66**, 486–501.
32   O. S. Smart, T. . Womach, A. Sharff, C. Flensburg, P. Keller, W. Paciorek, C. Vonrhein and G. Bricogne. Grade, v.1.2.9. Global Phasing Ltd., Cambridge, UK.
33   C. Vonrhein, C. Flensburg, P. Keller, A. Sharff, O. Smart, W. Paciorek, T. Womack and G. Bricogne, *Acta Crystallogr. Sect. D Biol. Crystallogr.*, 2011, **67**, 293–302.
34   I. J. Tickle, C. Flensburg, P. Keller, W. . Paciorek, A. Sharff, C. Vonrhein and G. Bricogne, 2018.
35   M. Machius, J. L. Chuang, R. M. Wynn, D. R. Tomchick and D. T. Chuang, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 11218–11223.
36   G. N. Murshudov, A. A. Vagin and E. J. Dodson, *Acta Crystallogr. D. Biol. Crystallogr.*, 1997, **53**, 240–55.
37   P. Emsley, B. Lohkamp, W. G. Scott and K. Cowtan, *Acta Crystallogr. D. Biol. Crystallogr.*, 2010, **66**, 486–501.
38   G. Bricogne, E. Blanc, M. Brandl, C. Flensburg, P. Keller, W. Paciorek, P. Roversi, A. Sharff, O. S. Smart, C. Vonrhein and T. O. Womack. BUSTER v.2.11.7., Global Phasing Ltd., Cambridge, UK
39   AFITT-CL 2.4.1.2: OpenEye Scientific Software, Santa Fe, NM.
40   N. Tzarum, Y. Eisenberg-Domovich, J. J. Gills, P. A. Dennis and O. Livnah, *J. Mol. Biol.*, 2012, **424**, 339–353.
41   J. J. P. Perry, R. M. Harris, D. Moiani, A. J. Olson and J. A. Tainer, *J. Mol. Biol.*, 2009, **391**, 1–11.
42   G. F. De Nicola, E. D. Martin, A. Chaikuad, R. Bassi, J. Clark, L. Martino, S. Verma, P. Sicard, R. Tata, R. A. Atkinson, S. Knapp, M. R. Conte and M. S. Marber, *Nat. Struct. Mol. Biol.*, 2013, **20**, 1182–1190.
43   T. A. Springer and M. L. Dustin, *Curr. Opin. Cell Biol.*, 2012, **24**, 107–115.
44   N. C. Nicolaides, Z. Galata, T. Kino, G. P. Chrousos and E. Charmandari, *Steroids*, 2010, 75, 1–12.
45   S. C. Tso, X. Qi, W. J. Gui, J. L. Chuang, L. K. Morlock, A. L. Wallace, K. Ahmed, S. Laxman, P. M. Campeau, B. H. Lee, S. M. Hutson, B. P. Tu, N. S. Williams, U. K. Tambar, R. M. Wynn and D. T. Chuang, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 9728–9733.

46   C. L. Quinlan, S. E. Kaiser, B. Bolaños, D. Nowlin, R. Grantner, S. Karlicek-Bryant, J. L. Feng, S. Jenkinson, K. Freeman-Cook, S. G. Dann, X. Wang, P. A. Wells, V. R. Fantin, A. E. Stewart and S. K. Grant, *Nat. Chem. Biol.*, 2017, **13**, 785–792.

47   R. A. Laskowski, M. W. MacArthur, D. S. Moss and J. M. Thornton, *J. Appl. Crystallogr.*, 1993, **26**, 283–291.