

RESEARCH ARTICLE

Modeling conformational flexibility of kinases in inactive states

Dominik Schwarz^{1,2} | Benjamin Merget¹ | Charlotte Deane²  | Simone Fulle¹ ¹BioMed X Innovation Center, Heidelberg, Germany²Department of Statistics, University of Oxford, Oxford, UK**Correspondence**Simone Fulle, BioMed X Innovation Center, Im Neuenheimer Feld 515, Heidelberg, Germany
Email: simone.fulle@gmail.com**Present address**

Benjamin Merget, BASF SE, Computational Chemistry, Carl-Bosch-Str. 38, 67063 Ludwigshafen, Germany.

Simone Fulle, Novo Nordisk A/S, Modelling and Predictive Technologies, 2760 Måløv, Denmark.

Abstract

Kinase structures in the inactive “DFG-out” state provide a wealth of druggable binding site variants. The conformational plasticity of this state can be mainly described by different conformations of binding site-forming elements such as DFG motif, A-loop, P-loop, and α C-helix. Compared to DFG-in structures, DFG-out structures are largely underrepresented in the Protein Data Bank (PDB). Thus, structure-based drug design efforts for DFG-out inhibitors may benefit from an efficient approach to generate an ensemble of DFG-out structures. Accordingly, the presented modeling pipeline systematically generates homology models of kinases in several DFG-out conformations based on a sophisticated creation of template structures that represent the major states of the flexible structural elements. Eighteen template classes were initially selected from all available kinase structures in the PDB and subsequently employed for modeling the entire kinome in different DFG-out variants by fusing individual structural elements to multiple chimeric template structures. Molecular dynamics simulations revealed that conformational transitions between the different DFG-out states generally do not occur within trajectories of a few hundred nanoseconds length. This underlines the benefits of the presented homology modeling pipeline to generate relevant conformations of “DFG-out” kinase structures for subsequent in silico screening or binding site analysis studies.

KEYWORDS

conformational ensembles, DFG-out, drug design, homology modeling, kinases

1 | INTRODUCTION

Protein kinases are key players of cellular signal transduction.¹ Since misregulation of kinases or mutations occur in several diseases, such as cancer and inflammation, human kinases comprise a large group of potential drug targets.² The catalytic domain of kinase structures has a bilobal architecture, with an N-(terminal) lobe consisting of β -strands and one α -helix (called α C-helix), and a C-lobe consisting of mainly α -helices. Connected by a flexible hinge region, the two lobes form a cleft with an evolutionarily conserved ATP-binding pocket and the catalytic center.³ The C-terminal domain consists of a flexible activation (A)-loop, typically 20-30 amino acids in length and marked by a

conserved Asp-Phe-Gly (DFG) motif at the beginning of the A-loop.⁴ The A-loop can undergo large conformational changes, thereby controlling the catalytic activity and access to the substrate-binding pocket. The conformational change is often accompanied by a specific orientation of the sidechains of the DFG motif, which leads to binding of different classes of inhibitors. In the catalytically active “DFG-in” state, two Mg^{2+} -ions are positioned by the DFG aspartate and the catalytic loop, which coordinate the transfer of ATP phosphate groups.⁵ Compounds that preferentially bind to the DFG-in conformation are called type I inhibitors. In turn, in the inactive “DFG-out” conformation, the side chains of Asp and Phe of the DFG motif are flipped compared to the DFG-in state, leading the Asp pointing away from the binding site.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 The Authors. *Proteins: Structure, Function, and Bioinformatics* published by Wiley Periodicals, Inc.

This prevents Mg^{2+} -coordination and, hence, catalytic activity.³ The DFG-out conformation furthermore creates a hydrophobic pocket adjacent to the ATP binding site that can be targeted by the so-called type II inhibitors. Another important structural element-determining catalytic activity is the position of the α C-helix, which consists of a conserved Glu residue and forms in the active state a crucial salt bridge with Lys of the N-lobe β 3-strand but is usually moved outward in the inactive (DFG-out) state (Figure 1).^{5,6} Overall, there are a plethora of inactive conformations with varying A-loop, DFG, and α C-helix orientations. Furthermore, the P-loop (also called G-loop) is also highly flexible and can stack down onto the ligand, thereby creating a more buried cavity and stabilizing interactions with the ligand. A detailed analysis and comprehensive classification of the sampled conformation of kinases in the Protein Data Bank (PDB) can be found in a study by Möbitz.⁵

Both states, DFG-in and DFG-out, have been successfully employed for the design of approved kinase inhibitors^{7,8} and have promises and challenges for the design of novel kinase inhibitors.^{9,10} For instance, due to the high structural conservation of the active site, developing selective kinase inhibitors might be essential to reduce undesired off-target activities.¹¹⁻¹³ Structure-based design efforts rely on structures in atomic detail and benefit from the presence of structural ensembles. Of all kinase structures available in the PDB, the "DFG-out" conformation covers only ~15% of all typical human kinases, which is much lower than the ~45% covered in the DFG-in state. This hinders classical homology modeling based on single template structures and already resulted in several attempts to model the DFG-out states via more sophisticated methods. For instance, the DOLPHIN (deletion-of-loop Asp-Phe-Gly-in) method¹⁴ removes the DFG-Phe and four adjacent residues from DFG-in structures in order to resemble the hydrophobic pocket occupied by type II inhibitors. The activation-loop remodeling method (ALRM) method¹⁵ accounts for variability in the A-loop and an N-lobe rotation, while the DFGmodel approach¹⁶ generates multiple models for a single kinase based on a selection of representative template structures considering mainly the relative positions of the N/C-lobes. However, none of these three approaches considers the structural diversity of all structural elements characterizing inactive ("DFG-out") structures. On the other side, large-scale approaches, such

as employing all available kinase catalytic domain structures to construct models of the human tyrosine kinase family,¹⁷ are very computationally demanding and may only shift the decision which structures to use for subsequent analysis.

Here, we describe a systematic generation of kinase structures in different DFG-out states by generating multiple possible combinations of the A-loop, P-loop, and α C-helix. Thus, the presented homology modeling approach accounts for the flexibility of the inactive state of kinases and generates a representative subset of "DFG-out" conformations for binding site analysis or screening studies.

2 | MATERIALS AND METHODS

2.1 | Structural classification

In order to group the relevant "DFG-out" conformations (Figure 1), all available PDB structures of human kinase domains (dated 08/09/2016) were classified based on geometrical features as described below. All residue IDs refer to the numbering of the PKAC α structure with the PDB code 1ATP.

2.1.1 | DFG classification

The conformation of the DFG motif was initially determined to select only "DFG-out" structures as templates of the kinase N-lobes. The employed DFG classification is based on cross products of vectors of four atoms of the DFG motif as described by Xu et al.¹⁵ All DFG-out structures in the PDB were further characterized based on structural variations in A-loop, P-loop, and α C-helix.

2.1.2 | A-loop classification

Kinase structures in the PDB were previously clustered by Möbitz into 14 major conformations based on pseudotorsional angles between four consecutive C_{α} atoms around the DFG motif.⁵ Three main clusters of A-loop conformations for "DFG-out" structures were initially defined based on visual inspection and called in analogy to Möbitz "closed

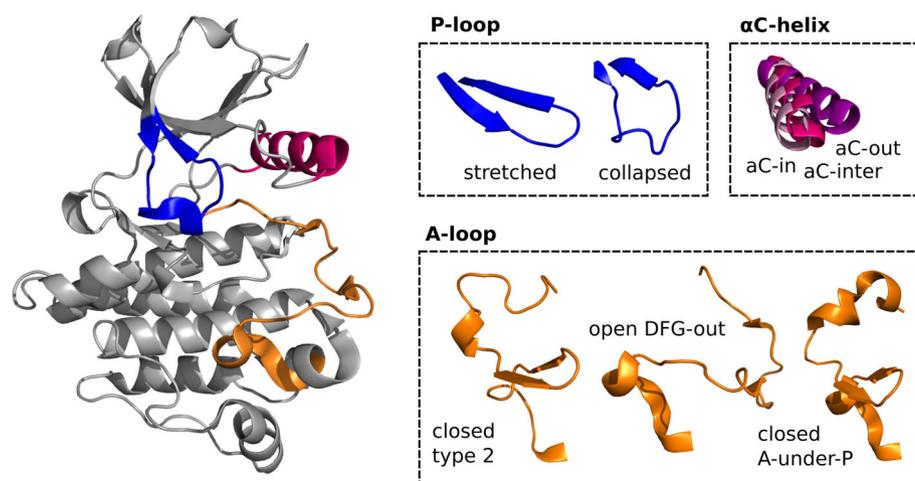


FIGURE 1 Structural elements of kinases that are considered for the generation of DFG-out ensembles. P-loop: stretched and collapsed; α C-helix: α C{-in, -inter, -out}; A-loop: closed {type 2, A-under-P}, and open DFG-out. Fusing these structural states results into 18 chimeric template structures for homology modeling

type 2," "open DFG-out," and "closed A-under-P" throughout the article. They were identified by employing pseudotorsional angles $\chi_{\text{DFG}\{-1, \text{D}\}}$ and $\chi_{\text{DFG}\{\text{F}, \text{G}\}}$ (already used by Möbitz⁵) and a $\text{C}_\alpha\text{-C}_\alpha$ distance criterion between a residue in proximity to the His-Arg-Asp (HRD) motif in the upper right part of the C-lobe (ie, HRD₋₄; ID 160) and a residue in the A-loop (ie, DFG₊₃; ID 189) in order to obtain more homogeneous clusters (Table S1).

P-loop classification is based on the evaluation of four structural features: two backbone dihedrals of the residues just before (ID 49; $\psi_{\text{G-motif-1}}$) and after (ID 56; $\psi_{\text{G-motif+1}}$) the GxGxPhiG motif, a pseudotorsional angle between the C_α atoms of the four residues following *Phi* (IDs 55-58; $\chi_{\text{G-motif}\{+1, +2\}}$), and a $\text{C}_\alpha\text{-C}_\alpha$ distance between the *Phi* residue (ID 54) and HRD₊₄ (ID 170). If at least three of these conditions are fulfilled (see Table S2), the P-loop is classified as either "collapsed" or "stretched." The four features were extracted from a feature importance analysis, employing a random forest classifier trained on manual P-loop class annotations (see Figure S4 for further details of the classifier development).

α C-helix classification is done following the rules described by Brooijmans et al⁶ that employs the minimal distance between the catalytic Lys (ID 72, atom NZ) and α C-helix's Glu (ID 91, OE1, or OE2) to differentiate between α C-in ($d \leq 4 \text{ \AA}$) and α C-out ($d \geq 8.5 \text{ \AA}$) conformations. For distances in between, the α C-helix's Glu dihedral χ_{i+1} is considered (ie, α C-inter: if angle $\leq 100^\circ$; α C-in: otherwise).

2.2 | Homology modeling

Homology modeling was performed with the YASARA program,¹⁸ employing pre-prepared template structures and alignments as well as the following parameters: the number of templates to use: 1; the number of ambiguous alignments to consider per template: 1; the number of samples to try per loop: 25; and the maximum number of unaligned terminal residues to model: 10.

TABLE 1 N-lobe templates for DFG-out structures with certain P-loop/ α C-helix combinations

| Structural class | Number of | | Selected structure | | |
|------------------|------------------|----------------|--------------------|---------------|--------------|
| | α C-helix | PDB structures | Unique kinases | PDB code | Kinase |
| Collapsed | α C-in | 33 | 6 | 4QQ5; chain A | FGFR4 (TK) |
| | α C-inter | 4 | 3 | 2G2H; chain B | ABL1 (TK) |
| | α C-out | 5 | 4 | 5HX6; chain A | RIPK1 (TKL) |
| Stretched | α C-in | 246 | 41 | 3VHK; chain A | KDR (TK) |
| | α C-inter | 39 | 23 | 4PMM; chain A | TRKA (TK) |
| | α C-out | 51 | 15 | 2W5B; chain A | NEK2 (Other) |

TABLE 2 A-loop templates for DFG-out structures with certain A-loop conformations

| Structural class | Number of | | Selected structure | | |
|------------------|-----------|----------------|--------------------|---------------|-----------|
| | A-loop | PDB structures | Unique kinases | PDB code | Kinase |
| Closed type 2 | | 37 | 10 | 3V5Q; chain A | TRKC (TK) |
| Open DFG-out | | 55 | 16 | 2HZI; chain A | ABL1 (TK) |
| Closed A-under-P | | 88 | 17 | 3BEA; chain A | FMS (TK) |

2.2.1 | Template construction

Template structures consisted of (rigid) C-lobes of a corresponding "DFG-in" structure (without the A-loop), an N-lobe of one of the six selected N-lobe structures, and an A-loop of one of the three selected A-loop structures (Tables 1 and 2). Full kinase domains of this "N-lobe" and "A-loop" structure representatives were structurally aligned to the C-lobe of "DFG-in" structures by only considering the C-lobe residues (without the A-loop). Then, all residues except for the desired ones of the respective structure were deleted and the remaining structural elements joined into one "chimeric template structure." Finally, a short energy minimization was performed to remove steric clashes.

2.2.2 | Target-to-template alignment

The alignment provided in Reference 5 by Möbitz is a manually curated multiple sequence alignment of nearly 500 kinases. The sequence parts of the chimeric template structures were aligned to the corresponding sequence parts of the same kinase in the Möbitz alignment using the pairwise2.align.globalmc alignment function in Biopython, joined into one sequence, and finally employed as template sequence. The canonical catalytic kinase domain sequence from UniProt was employed as target sequence in the modeling step and accordingly also aligned to the alignment (with the exception of sequences of the kinases MASTL, SgK494, NEK10, and MNK1 whose UniProt sequences were either lacking important sequence parts or contained unwanted insertions). Hence, their target sequences were taken from References 5 and 19. A list of UniProt IDs can be found on <http://www.kinhub.org/kinases.html>.

2.2.3 | Input structure selection

"DFG-in" input structures (for the C-lobe templates) were taken from an in-house "selected set" of kinase structures (ie, updated version of

the version described in Reference 13) that consists of manually selected high-quality DFG-in structures per kinase. The following criteria were considered for the selection of the “selected set”: high resolution, DFG state (in/out), and structural completeness (especially in terms of having resolved A- and P-loops), and favoring structures bound to ATP or ATP analogues. If a kinase was not present in the “selected set,” a homology model was generated employing the closest structure in the “selected set” as a template.

2.2.4 | Template structures

A total of 18 chimeric template structures were generated that represent the structural variation of the DFG-out state and consist of two major P-loop (ie, collapsed and stretched), three α C-helices (ie, α C-[in, inter, out]), and three A-loop conformations (ie, closed type 2, open DFG-out, and closed A-under-P). The selected N-lobe and A-loop template structures as well as one exemplary set of generated homology models (ie, for ABL1) can be obtained from <https://github.com/Team-SKI/Publications>.

2.3 | Molecular dynamics simulations

2.3.1 | Structure preparation and molecular dynamics parameters

ABL1 homology models of types 8, 12, and 16 were protonated, parameterized with the ff14SB force field, charge neutralized with Na^+ ions, and solvated with TIP3P water using the tLeAP module of AmberTools16.²⁰ All simulations were carried out with NAMD,²¹ using 2 fs simulation time steps and the SHAKE²² algorithm to restrain all bonds to hydrogen atoms. The nonbonded energy calculation cut-off distance was set to 12 Å, switching distance to 10 Å, and pair list distance to 13.5 Å with 10 steps per cycle. Electrostatic interactions were calculated with particle mesh Ewald (PME) method²³ and temperature regulated with Langevin dynamics.

2.3.2 | Equilibration

10 000 steps of minimization were followed by a 1 ns equilibration in the NVT ensemble which heats from 100 K to 300 K by increasing the temperature by 1 K every 2.5 ps. Initial 0.5 kcal/(mol Å²) constraints were gradually turned off after 100 ps until 500 ps, by 0.1% every 4 ps, ending the equilibration with another 500 ps of unconstrained simulation. 200 ns production runs were carried out in the NPT ensemble at 1.01325 bar with the Langevin piston Nosé-Hoover barostat.^{24,25}

2.3.3 | Trajectory analysis

Coordinates were taken at regular intervals of 10 ps during the simulation. To assess the structural variability of the obtained homology models, backbone root-mean-square deviations (RMSDs) to the energy-minimized starting structures were calculated (Figures S2 and S3). VMD 1.9.2 was used for aligning the trajectories to the

corresponding starting structures and calculating the RMSD values. The statistical framework R was used for analysis and visualization.

2.4 | Kinome tree plots

All kinome tree plots were generated via KinMap (www.kinhub.org/kinmap/).²⁶

3 | RESULTS

3.1 | Structural classification

The generation of an “DFG-out” ensemble is based on the usage of selected structures comprising the major different states of flexible structural elements of the kinase domain (Figure 1) and a subsequent fusion of these elements to multiple chimeric template structures. The initially selected structures were obtained by classifying all “DFG-out” structures in the PDB with respect to the possible configurations of the A-loop, P-loop, and α C-helix (see the Materials and Methods section). The analysis revealed that three major structural variants of the A-loop, two of the P-loop, and three of the α C-helix exist (Tables 1 and 2).

Furthermore, since seven variants occurred in at least six out of the eight kinase groups, and the remaining variant (A-loop: closed type 2) in three, we assumed that they are all distributed across the kinome (Data S1, Figure S1). Noteworthy, a kinase profiling study of presumable type II (DFG-out) inhibitors revealed that more than 200 kinases, covering all branches of the kinome, were targeted by this set of inhibitors.⁹ This suggests that DFG-out conformations are commonly sampled in the majority of kinases, and thus, seem not to be a unique feature of a particular subset of kinases. Thus, we assumed in the modeling pipeline that all kinases are able to adopt all possible combinations of the A-loop, P-loop, and α C-helix.

Compared with the DFG-in state, the N-lobe in the DFG-out state is rotating relative to the C-lobe. Hence, we aimed for identifying one template each for the N-lobe and C-lobe as well as for the A-loop. The C-lobe (excluding the A-loop) can be considered to be rigid (ie, it is not affected by the DFG conformation) and was taken from either a manually selected high-quality DFG-in structure of the respective kinase or a generated homology model (see the Materials and Methods section for more details as well as Reference 13). The N-lobe can sample two major P-loop and three major α C-helix conformations (ie, P-loop: collapsed and stretched; α C-helix: α C-[in, inter, out]), resulting in six possible combinations. The α C-helix can be classified depending on whether a salt bridge can be formed with the catalytic Lys of the N-lobe (α C-in; present in 80% of active but only in 36% of inactive conformations⁵) or not, either due to a rotation of the α C-helix (α C-out) or despite no larger conformational changes of the α C-helix (α C-inter). The final selection of structures for the N-lobe part is listed in Table 1 and only includes the structures with a “DFG-out” motif, which passed geometrical criteria for these classes (see the Materials and Methods section), and contained no ambiguous or missing residues in the N-lobe.

The A-loop samples three respective clusters in DFG-out structures (ie, closed type 2, open DFG-out, and closed A-under-P). “Closed

type 2" is the classical DFG-out conformation for type II inhibitors, where the side chains of Asp and Phe of the DFG motif are flipped compared with the DFG-in state, and the A-loop is in a closed folded conformation. In contrast, the A-loop is, similar to DFG-in structures, in an unfolded conformation in the "open DFG-out" conformation that opens up the active site and is compatible with binding of type I inhibitors. "Closed A-under-P" conformations contain again A-loops in a closed conformation, but which form stabilizing stacking interactions with the P-loop. Please note that while Möbitz⁵ differentiated further between characteristic A-under-P conformations, we assigned them all to one structural class. An extended A-loop classifier (see the Materials and Methods section) was employed to cluster all kinase structures in the DFG-out state into these A-loop conformational classes, whereas the final selection for the A-loop templates also considered the completeness of the loop (Table 2).

3.2 | Homology modeling pipeline

Examples of the explicitly considered flexible structural elements of kinases as well as the homology modeling pipeline are displayed in Figures 1 and 2. The selected N-lobe (ie, α C-helix and P-loop) and A-

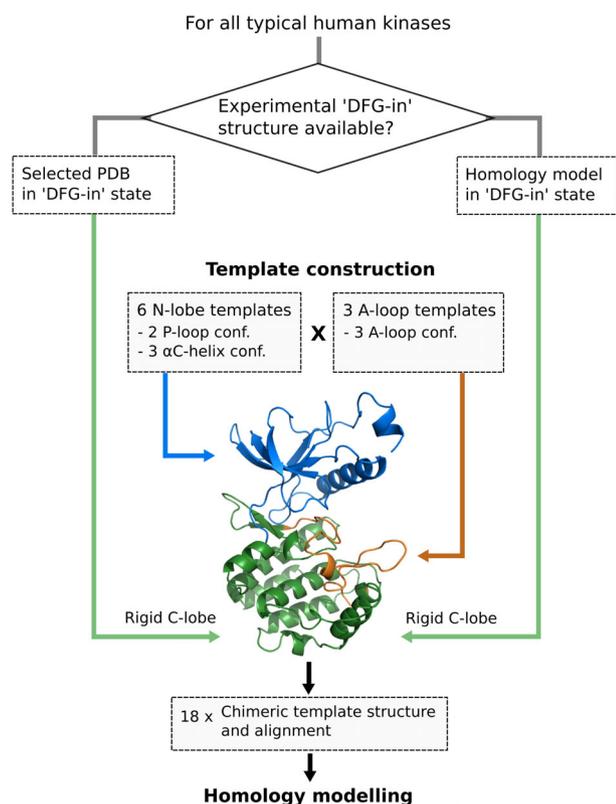


FIGURE 2 Scheme of homology modeling pipeline for modeling the entire kinase in different DFG-out conformations. Template structures are constituted by a C-lobe of a corresponding DFG-in structure (green), an N-lobe of one of the six selected N-lobe template structures (blue), and an A-loop of one of the three selected A-loop templates (orange), resulting into 18 chimeric template structures per kinase [Color figure can be viewed at wileyonlinelibrary.com]

loop variants are used in all possible combinations together with one C-lobe structure, forming a total of 18 chimeric templates. A target-to-template alignment was created by using the respective sequences of the target kinase and template segments of a manually curated multiple sequence alignment provided in Reference 5 (see the Materials and Methods section).

3.2.1 | Homology modeling of kinases in the DFG-out state

Almost all typical human kinases could successfully be modeled in different structural states. Z-scores, which estimate the overall quality of a model, indicate that all groups beside the "Other" group have a good quality with respect to dihedrals and packing parameters (Table 3). The best results were obtained for the TK group (median -0.54 ± 0.32), the lowest Z-scores for the "Other" group (-1.58 ± 0.89), while the remaining six groups have median values around -1 . Individual Z-scores of all obtained homology models, ordered by groups, are depicted in Figure 3. Models of the CK1, TK, and TKL groups are in general of good quality as only a few models obtained Z-scores below -2 . The AGC, CAMK, CMGC, and STE groups possess individual outlier kinases that score below -2 (note that the vertical series of adjacent data points usually display the models of a single target kinase). Visual inspection of randomly selected structures did not reveal any conspicuous features of these structures, such as unexpected folds or clashes. The distribution of Z-scores of the "Other" group is much more scattered with scores ranging from -4.5 to about 0 , presumably displaying the heterogeneity of this kinase group.

Out of the available PDB structures, ~ 160 structures could be assigned classes for all three flexible elements and, thus, provide a valuable source for comparison with the corresponding homology models. Matching PDB structures were found for 13 of the 18 model types. Visual comparison of these structural pairs let to a manual

TABLE 3 Summary of generated homology models per kinase group

| Kinase group | Z-score ^a | Number of missing kinases |
|--------------|----------------------|---------------------------|
| AGC | -0.87 ± 0.33 | 4/63 |
| CAMK | -0.95 ± 0.39 | 4/82 |
| CK1 | -0.80 ± 0.33 | 0/12 |
| CMGC | -1.00 ± 0.34 | 0/63 |
| Other | -1.58 ± 0.89 | 8/82 |
| STE | -1.07 ± 0.41 | 1/48 |
| TK | -0.54 ± 0.32 | 1/94 |
| TKL | -1.06 ± 0.44 | 1/43 |

^aMedian and median absolute deviation of Z-scores for each kinase group, as calculated by the YASARA program. Z-scores describe how many standard deviations the model quality differs from typical high-resolution X-ray structures. Scores above 0 are considered to be "optimal," between 0 and -1 "good," between -1 and -2 "satisfactory," between -2 and -3 "poor," and below -3 "bad."

setting of a backbone RMSD cut-off of 0.7 Å to declare matching pairs as structurally similar or different. Structural pairs below this threshold generally displayed not only a “classification match” (according to A-loop, P-loop, and α C-helix classifiers based on angle and distance criteria) but also a “visual match” (similar shapes of the flexible features). Based on this threshold, at least one generated

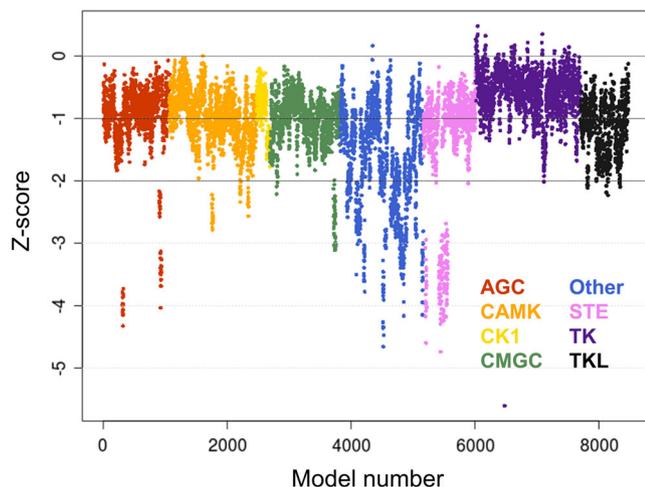


FIGURE 3 Distribution of Z-scores obtained for the generated homology models. Each kinase is represented by 18 models and color coded according to the respective kinase group (ie, AGC, red; CAMK, orange; CK1, yellow; CMGC, green; Other, blue; STE, purple; TK, violet; and TKL, black) [Color figure can be viewed at wileyonlinelibrary.com]

homology model was very similar to 10 of the 18 model types of experimental structures of which most were also, in general, similar in the course of respective molecular dynamics (MD) trajectories (Table 4). No matching PDB structures could be found for 5 model types consisting all of collapsed P-loop conformations. According to the analysis of available PDB structures, this P-loop conformation might prefer an A-loop in the open DFG-out state compared with closed type 2 and A-under-P conformations. Overall, the structural comparisons with existing PDB structures underline the value of the presented ensemble generation pipeline to generate scarcely populated conformations for further analysis and calculations.

3.3 | Probing structural variability/invariability with MD simulations

MD simulations of 200 ns length were run to assess the general structural variability of the generated homology models. ABL1 homology models of types 8, 12, and 16 were selected to include all three A-loop conformations (Figure 4; Table 5). The variability of the two P-loop conformations (“stretched” and “collapsed”) was also assessed as well as the two most different states of the α C-helix (“ α C-in” and “ α C-out”).

Visual inspection of the trajectories together with backbone RMSD calculations with respect to their energy-minimized starting structures (Table 5; Figure S2) indicates that ABL1 HM 12 is overall stable (ie, invariable) along the trajectory (mean RMSD: 2.26 ± 0.25), while small conformational changes occur during the HM 8 (mean RMSD: 2.64 ± 0.55) and HM 16 (mean RMSD: 3.19 ± 0.34).

TABLE 4 RMSD values of matching PDB structures and homology models

| HM model type | A-loop | P-loop | α C-helix | Matching PDB structures ^a | Mean RMSD | Minimal RMSD ^b |
|---------------|------------------|-----------|------------------|--------------------------------------|-----------------|---------------------------|
| HM 1 | Closed type 2 | Collapsed | α C-in | 17 | 0.59 ± 0.05 | 0.5 (ABL1, 2E2B_B) |
| HM 2 | Open DFG-out | Collapsed | α C-in | 7 | 0.53 ± 0.05 | 0.5 (ABL1, 2HZI_A) |
| HM 3 | Closed A-under-P | Collapsed | α C-in | 0 | - | - |
| HM 4 | Closed type 2 | Collapsed | α C-inter | 0 | - | - |
| HM 5 | Open DFG-out | Collapsed | α C-inter | 3 | 0.51 ± 0.06 | 0.5 (ABL1, 2G2H_B) |
| HM 6 | Closed A-under-P | Collapsed | α C-inter | 0 | - | - |
| HM 7 | Closed type 2 | Collapsed | α C-out | 0 | - | - |
| HM 8 | Open DFG-out | Collapsed | α C-out | 3 | 1.28 ± 0.32 | 1.1 (LIMK2, 4TPT_B) |
| HM 9 | Closed A-under-P | Collapsed | α C-out | 0 | - | - |
| HM 10 | Closed type 2 | Stretched | α C-in | 11 | 1.00 ± 0.19 | 0.6 (FGFR4, 4TYJ_A) |
| HM 11 | Open DFG-out | Stretched | α C-in | 14 | 0.73 ± 0.17 | 0.5 (ABL1, 3UE4_B) |
| HM 12 | Closed A-under-P | Stretched | α C-in | 69 | 0.70 ± 0.18 | 0.4 (KDR, 3VO3_A) |
| HM 13 | Closed type 2 | Stretched | α C-inter | 2 | 1.18 ± 0.02 | 1.2 (AMPKa2, 2YZA_A) |
| HM 14 | Open DFG-out | Stretched | α C-inter | 6 | 0.70 ± 0.22 | 0.5 (MAP3K5, 2CLQ_A) |
| HM 15 | Closed A-under-P | Stretched | α C-inter | 4 | 0.79 ± 0.09 | 0.7 (BRAF, 4R5Y_B) |
| HM 16 | Closed type 2 | Stretched | α C-out | 2 | 1.00 ± 0.00 | 1.0 (CDK6, 1G3N_A) |
| HM 17 | Open DFG-out | Stretched | α C-out | 9 | 1.23 ± 0.64 | 0.5 (ABL1, 4YC8_B) |
| HM 18 | Closed A-under-P | Stretched | α C-out | 11 | 0.89 ± 0.09 | 0.7 (LOK, 4AOT_B) |

^aRMSD cut-off for declaring a matching PDB structure/homology model pair as structurally similar was set to 0.7 Å.

^bCorresponding kinase is given in brackets.

FIGURE 4 Homology models of ABL1 that were used as starting structures of molecular dynamics simulations. P-loops are colored in blue, α C-helices in pink, and A-loops in orange. The A-loop conformation in model type HM 8 is “open DFG-out,” in HM 12 “closed A-under-P,” and in HM 18 “closed type 2” [Color figure can be viewed at wileyonlinelibrary.com]

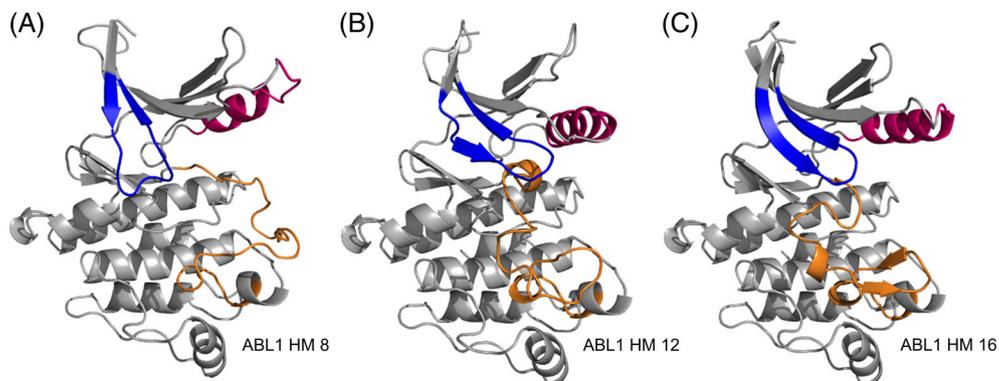


TABLE 5 Homology models chosen as starting structures for MD simulations

| Homology model | A-loop class | P-loop class | α C-helix class | Mean RMSD |
|----------------|------------------|--------------|------------------------|-----------------|
| ABL1–HM 8 | Open DFG-out | Collapsed | α C-out | 2.64 \pm 0.55 |
| ABL1–HM 12 | Closed A-under-P | Stretched | α C-in | 2.26 \pm 0.25 |
| ABL1–HM 16 | Closed type 2 | Stretched | α C-out | 3.19 \pm 0.34 |

The structural classifications refer to the chimeric template that was used to build the respective homology model. Mean backbone RMSD values were calculated over MD trajectories of 200 ns length with respect to their energy-minimized starting structure.

trajectories due to movements of the α C-helix and its connecting loop to the N-lobe's β -sheet. The conformational change sampled in the HM 8 trajectory might result from a suboptimal choice of the N-lobe template structure (5HX6; chain A), which possesses an unusually short α C-helix but was the only template structure matching this particular combination of P-loop and α C-helix at the time of the present study. Thus, another N-lobe template with a “collapsed” P-loop and an α C-helix in an “ α C-out” configuration might improve the quality of future homology modeling calculations. Apart from a turn formation of the α C-helix, the model structure is stable along the trajectory and does not undergo noteworthy conformational changes. In the case of the ABL1 HM 16 trajectory, a slight anticlockwise rotation of the N-lobe and subsequent electrostatic interactions between the connecting loop between α C-helix and β -sheet and the C-lobe occurs. After formation of these interactions, the RMSD increases very slowly throughout the simulation, and visual inspection shows no major structural changes.

The conformational changes sampled in the three MD trajectories were mainly related to movement of the α C-helix or its connecting elements, which is in line with the general mobile nature of the α C-helix.^{1,27} As expected, larger conformational changes (such as necessary for the A-loop) did not occur in any of the three MD trajectories, as they occur on the microsecond to millisecond timescale.²⁸ This underlines the benefit of the presented modeling pipeline to generate an ensemble of inactive (“DFG-out”) structures as well as the overall structural invariability of the generated homology models in the course of MD simulations of a few hundreds of nanoseconds length.

TABLE 6 Small docking set of FDA approved type II inhibitors

| Kinase | PDB code | Docked compounds |
|--------|-------------------------------|--------------------------------|
| ABL1 | 2HYY, 3CS9, 3OXZ ^a | Imatinib, nilotinib, ponatinib |
| KDR | 3WZE, 4AG8 ^b | Sorafenib, axitinib |

^aAll three ABL1 crystal structures are classified as A-loop: “closed type 2,” P-loop: “collapsed,” and α C-helix: “ α C-in” (which is equivalent to HM 1).

^bAll two KDR crystal structures are classified as A-loop: “closed A-under-P,” P-loop: “stretched,” and α C-helix: “ α C-in” (which is equivalent to HM 12).

An additional comparison of all three simulation trajectories against all other homology model types further supports this conclusion as, besides again the α C-helix configuration, no conformational transitions occur between the structural classes (Figure S3).

3.4 | Testing model ensembles for docking studies

Applications of the generated DFG-out ensembles are manifold, including their application for in silico screening and binding site analysis studies. To test whether our model ensembles are suited for docking studies, we selected FDA-approved type II inhibitors of the two kinases ABL1 and KDR (five cases in total; Table 6) and docked them into the generated DFG-out ensembles for those two kinases. Encouragingly, docking poses with good agreement to the crystal structure (ie, RMSD <2.2 Å) were found within the top five docking poses for four out of the five cases (Tables S4 and S5), and at least in the case of the ABL1 structure many of the corresponding receptor models have the same structural features /classification as the respective crystal structure (see Data S1 for more details).

For the three ABL1 cases, the respective crystal structures are all classified as A-loop: “closed type 2,” P-loop: “collapsed,” and α C-helix: “ α C-in” (which is equivalent to HM 1). For two of the respective docking cases (imatinib and nilotinib), a good docking pose and the correct model class were found within the top five docking poses, and for all three cases all receptor models of the top poses have at least a closed A-loop conformation. For the two KDR cases, the respective crystal structures are both classified as A-loop: “closed A-under-P,” P-loop: “stretched,” and α C-helix: “ α C-in” (which is equivalent to HM 12). For the case of sorafenib, a good docking pose was obtained for the

“closed type 2” model, which is similar in shape to the “closed A-under-P” conformation and three out of the top five selected structures have a closed A-loop conformation. In the case of axitinib, a matching pose was found in the top five docking solutions but the respective receptor model has a different A-loop orientation from the crystal structure. Overall, the results imply that the generated ensembles of homology models can be employed for docking calculations, especially when the top poses are found for the same receptor classification state.

4 | DISCUSSION

4.1 | Inactive “DFG-out” states—accessible for every kinase?

One fundamental question of kinome-wide modeling of DFG-out conformations is whether every kinase or only a subset of the kinome is capable of adopting the DFG-out state.⁹ The factors modulating DFG conformations still remain poorly understood. However, several aspects suggest that at least most kinases can adopt every major conformation at a conformational penalty.^{5,29,30} As mentioned earlier, a large ligand-based profiling study showed that many more kinases compared to those that have a “DFG-out” structure in the PDB can be targeted by type II inhibitors.⁹ This speaks in favor of most kinases being able to adopt the “DFG-out” state and might also imply that many of the so far untargeted kinases could potentially be targeted by type II inhibitors. Furthermore, also kinases that were thought to have an inaccessible DFG-out conformation, such as CDK2,³¹ can bind type II inhibitors and undergo the DFG-out transition.

Another crucial issue of modeling “DFG-out” kinases is not only the question of whether kinases generally may adapt a “DFG-out” state, but also how many different ones. It was shown that for instance ABL1 is able to adapt multiple inactive “DFG-out” conformations³² and that BTK can sample kinome-wide crystallographically observed states.³³ Energetic accessibility of the different states might be shifted by multiple factors, such as the phosphorylation state of the A-loop,¹⁶ binding of allosteric regulators,³⁴ or induction by inhibitors.⁵ This emphasizes the importance and validity of modeling different “DFG-out” states that have not been observed in crystal structures yet. Most likely not all of the 18 combinations of structural features can be adopted by all kinases but exceptions to that can be accepted while the general idea is valid, especially when it comes to large-scale modeling and analysis studies.

4.2 | Modeling conformational plasticity of DFG-out structures

Previous “DFG-out” modeling approaches consist of the ALRM,¹⁵ DOLPHIN,¹⁴ and DFGmodel¹⁶ approaches. The ALRM approach is addressing the conformational flexibility of “DFG-out” states by generating a large number of possible A-loop conformations and subsequently filtering the generated structures based on the space of the active-site cleft.¹⁵ Furthermore, the N-lobes were rotated about a

predefined axis. Thus, ALRM accounts for the variance in the A-loop and an N-lobe rotation but not for the other N-lobe variations, such as different conformations of the P-loop and α C-helix. The presented approach accounts for two different P-loop conformations and three different α C-helix states but models a lower number of different A-loop conformations. However, modeling the entire accessible A-loop conformations might not be necessary for hit identification tasks of small molecules as only the first part, including the DFG motif, and some subsequent residues affect ligand binding; modeling more A-loop conformations might be of relevance when it comes to the analysis and modeling of protein substrate binding though.⁴

The DFGmodel approach,¹⁶ as well as DOLPHIN method,¹⁴ therefore excludes A-loop residues after the DFG motif from the modeling procedure (ie, DFGmodel: DFG_F and the next four residues; DOLPHIN: beyond DFG motif), resulting nevertheless in structures with high practical value for virtual screening. However, also residues after the DFG motif can be involved in forming the binding site (eg, the A-under-P structure 3BEA). Furthermore, the presence of the A-loop might be required for subsequent MD studies.

Similar to the present study, modeling via the DFGmodel approach¹⁶ is also done based on an ensemble of 18 different template structures. However, the structures were selected based on large variations in the relative position of the N/C lobes and not manually created as done in the present approach. Thus, DFGmodel employs different criteria for the selection/creation of the template structures/modeling of the structural variety of inactive DFG-out states. Noteworthy, different templates were selected for tyrosine kinases and serine/threonine kinases. Differentiating between different kinase groups might be an interesting improvement of the presented template selection procedure and modeling pipeline, especially when more PDB structures with varying structural elements became available.

5 | CONCLUSION

The here presented modeling pipeline systematically generates homology models of kinases in the “DFG-out” state and, thereby, efficiently samples the major conformations of potential value for drug design efforts. The generated structures can not only be employed for docking calculations but can also provide informative insights into selectivity-determining features, which might be only addressable in scarcely populated conformations.³⁵ The basis of the presented pipeline forms a thorough selection of template structures, which were identified after classifying all available kinase structures in the PDB by rules either already described in the literature or newly developed. A systematically diverse set of template structures was identified and employed for the generation of homology models in different “DFG-out” conformations for all typical kinases in the human kinome. MD simulations further point to the value of the generated ensembles as conformational transitions of the flexible features of the kinase domains (except the α C-helix) did not occur within the sampled time-scale, underlining the potential of the presented ensemble generation pipeline for drug design efforts.

ACKNOWLEDGMENTS

We would like to thank Katra Kolsek and Guillaume Roellinger for preliminary work on kinase classification at BioMed X. Furthermore, we would like to acknowledge EPSRC grant EP/L016044/1 and the support of the German Academic Scholarship Foundation.

ORCID

Charlotte Deane  <https://orcid.org/0000-0003-1388-2252>

Simone Fulle  <https://orcid.org/0000-0002-7646-5889>

REFERENCES

- Taylor SS, Kornev AP. Protein kinases: evolution of dynamic regulatory proteins. *Trends Biochem Sci.* 2011;36(2):65-77.
- Cohen P. Protein kinases--the major drug targets of the twenty-first century? *Nat Rev Drug Discov.* 2002;1(4):309-315.
- Vijayan R, He P, Modi V, et al. Conformational analysis of the DFG-out kinase motif and biochemical profiling of structurally validated type II inhibitors. *J Med Chem.* 2015;58(1):466-479.
- Nolen B, Taylor S, Ghosh G. Regulation of protein kinases: controlling activity through activation segment conformation. *Mol Cell.* 2004;15(5):661-675.
- Möbitz H. The ABC of protein kinase conformations. *Biochim Biophys Acta.* 2015;1854(10):1555-1566.
- Brooijmans N, Chang Y-W, Mobilio D, Denny RA, Humblet C. An enriched structural kinase database to enable kinome-wide structure-based analyses and drug discovery. *Protein Sci.* 2010;19(4):763-774.
- Müller S, Chaikuad A, Gray NS, Knapp S. The ins and outs of selective kinase inhibitor development. *Nat Chem Biol.* 2015;11(11):818-821.
- Wu P, Nielsen TE, Clausen MH. Small-molecule kinase inhibitors: an analysis of FDA-approved drugs. *Drug Discov Today.* 2016;21(1):5-10.
- Zhao Z, Wu H, Wang L, et al. Exploration of type II binding mode: a privileged approach for kinase inhibitor focused drug discovery? *ACS Chem Biol.* 2014;9(6):1230-1241.
- Bosc N, Wroblowski B, Aci-Sèche S, Meyer C, Bonnet P. A proteomic analysis of human kinome: insight into discriminant conformation-dependent residues. *ACS Chem Biol.* 2015;10(12):2827-2840.
- Norman RA, Toader D, Ferguson AD. Structural approaches to obtain kinase selectivity. *Trends Pharmacol Sci.* 2012;33(5):273-278.
- Fulle S, Merget B, Turk S. Towards in silico design of selective kinase inhibitors. *Systembiology.eng.* 2017;11:40-43.
- Volkamer A, Eid S, Turk S, Jaeger S, Rippmann F, Fulle S. The pocketome of human kinases: prioritizing the ATP binding sites of (yet) untapped protein kinases for drug discovery. *J Chem Inf Model.* 2015;55(3):538-549.
- Kufareva I, Abagyan R. Type-II kinase inhibitor docking, screening, and profiling using modified structures of active kinase states. *J Med Chem.* 2008;51(24):7921-7932.
- Xu M, Yu L, Wan B, Yu L, Huang Q. Predicting inactive conformations of protein kinases using active structures: conformational selection of type-II inhibitors. *PLoS One.* 2011;6(7):e22644.
- Ung PM, Schlessinger A. DFGmodel: predicting protein kinase structures in inactive states for structure-based discovery of type-II inhibitors. *ACS Chem Biol.* 2015;10(1):269-278.
- Parton DL, Grinaway PB, Hanson SM, Beauchamp KA, Chodera JD. Ensembler: enabling high-throughput molecular simulations at the superfamily scale. *PLoS Comput Biol.* 2016;12:e1004728.
- Krieger, E., Vriend, G., and Spronk, C. YASARA--Yet Another Scientific Artificial Reality Application. <http://www.yasara.org/>. Accessed 1 March 2019.
- Ocasio CA, Rajasekaran MB, Walker S, et al. A first generation inhibitor of human Greatwall kinase, enabled by structural and functional characterisation of a minimal kinase domain construct. *Oncotarget.* 2016;7(44):71182-71197.
- Case DA, Betz RM, Botello-Smith W, et al. *Amber 2016*. San Francisco: University of California; 2016.
- Phillips JC, Braun R, Wang W, et al. Scalable molecular dynamics with NAMD. *J Comput Chem.* 2005;26(16):1781-1802.
- Ryckaert JP, Ciccotti G, Berendsen HJC. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J Comput Phys.* 1977;23(3):327-341.
- Darden T, York D, Pedersen L. Particle mesh Ewald: an $N * \log(N)$ method for Ewald sums in large systems. *J Chem Phys.* 1993;98(12):10089-10092.
- Martyna GJ, Tobias DJ, Klein ML. Constant pressure molecular dynamics algorithms. *J Chem Phys.* 1994;101(5):4177-4189.
- Feller SE, Zhang Y, Pastor RW, Brooks BR. Constant pressure molecular dynamics simulation: the Langevin piston method. *J Chem Phys.* 1995;103(11):4613-4621.
- Eid S, Turk S, Volkamer A, Rippmann F, Fulle S. KinMap: a web-based tool for interactive navigation through human kinome data. *BMC Bioinformatics.* 2017;18:16.
- Palmieri L, Rastelli G. α C helix displacement as a general approach for allosteric modulation of protein kinases. *Drug Discov Today.* 2013;18(7-8):407-414.
- Xiao Y, Liddle JC, Pardi A, Ahn NG. Dynamics of protein kinases: insights from nuclear magnetic resonance. *Acc Chem Res.* 2015;48(4):1106-1114.
- Ung P, Rahman R, Schlessinger A. Redefining the protein kinase conformational space with machine learning. *Cell Chem Biol.* 2018;25(7):916-924.
- Haldane A, Flynn W, He P, Vijayan R, Levy R. Structural propensities of kinase family proteins from a Potts model of residue co-variation. *Protein Sci.* 2016;25(8):1378-1384.
- Alexander LT, Möbitz H, Drueckes P, et al. Type II inhibitors targeting CDK2. *ACS Chem Biol.* 2015;10(9):2116-2125.
- Levinson NM, Kuchment O, Shen K, et al. A Src-like inactive conformation in the Abl tyrosine kinase domain. *PLoS Biol.* 2006;4(5):e144.
- Sultan MM, Denny RA, Unwalla R, Lovering F, Pande VS. Millisecond dynamics of BTK reveal kinome-wide conformational plasticity within the apo kinase domain. *Sci Rep.* 2017;7:15604. <https://doi.org/10.1038/s41598-017-10697-0>.
- Jura N, Zhang X, Endres NF, Seeliger MA, Schindler T, Kuriyan J. Catalytic control in the EGF receptor and its connection to general kinase regulatory mechanisms. *Mol Cell.* 2011;42(1):9-22.
- Guimarães CRW, Rai BK, Munchhof MJ, et al. Understanding the impact of the P-loop conformation on kinase selectivity. *J Chem Inf Model.* 2011;51(6):1199-1204.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Schwarz D, Merget B, Deane C, Fulle S. Modeling conformational flexibility of kinases in inactive states. *Proteins.* 2019;87:943-951. <https://doi.org/10.1002/prot.25756>