Exploring protein kinase conformation using swarm-enhanced sampling molecular dynamics (sesMD)
Alessio Atzori, Neil J Bruce, Kepa K Burusco, Berthold Wroblowski, Pascal Bonnet, and Richard A Bryce

J. Chem. Inf. Model., Just Accepted Manuscript • DOI: 10.1021/ci5003334 • Publication Date (Web): 01 Sep 2014

Downloaded from http://pubs.acs.org on September 2, 2014

Just Accepted

“Just Accepted” manuscripts have been peer-reviewed and accepted for publication. They are posted online prior to technical editing, formatting for publication and author proofing. The American Chemical Society provides “Just Accepted” as a free service to the research community to expedite the dissemination of scientific material as soon as possible after acceptance. “Just Accepted” manuscripts appear in full in PDF format accompanied by an HTML abstract. “Just Accepted” manuscripts have been fully peer reviewed, but should not be considered the official version of record. They are accessible to all readers and citable by the Digital Object Identifier (DOI®). “Just Accepted” is an optional service offered to authors. Therefore, the “Just Accepted” Web site may not include all articles that will be published in the journal. After a manuscript is technically edited and formatted, it will be removed from the “Just Accepted” Web site and published as an ASAP article. Note that technical editing may introduce minor changes to the manuscript text and/or graphics which could affect content, and all legal disclaimers and ethical guidelines that apply to the journal pertain. ACS cannot be held responsible for errors or consequences arising from the use of information contained in these “Just Accepted” manuscripts.
Exploring protein kinase conformation using swarm-enhanced sampling molecular dynamics (sesMD)

Alessio Atzori, Neil J. Bruce, Kepa K. Burusco, Berthold Wroblowski, Pascal Bonnet and Richard A. Bryce

a Manchester Pharmacy School, University of Manchester, Oxford Road, Manchester, M13 9PT, UK

b Janssen Research & Development, a division of Janssen Pharmaceutica N.V., Turnhoutseweg 30, 2340 Beerse, Belgium

c Structural Bioinformatics & Chemoinformatics, Institut de Chimie Organique et Analytique (ICOA), UMR CNRS-Université d’Orléans 7311, Université d’Orléans, Rue de Chartres, F-45067 Orléans cedex 02, France

‡ Current address: Heidelberg Institute for Theoretical Studies (HITS gGmbH), Schloss-Wolfsbrunnenweg 35, 69118 Heidelberg, Germany

¶ These authors contributed equally to this work.

*Author to whom all correspondence should be addressed:

Richard Bryce, Tel: +44 161 275 8345, Fax: +44 161 275 2396, E-mail r.a.bryce@manchester.ac.uk
Abstract

Protein plasticity, whilst often linked to biological function, also provides opportunities for rational design of selective and potent inhibitors of their function. The application of computational methods to predict concealed protein concavities is challenging, given that the motions involved can be significant and occur over long time scales. Here, we introduce the swarm-enhanced sampling molecular dynamics (sesMD) method as a tool to improve sampling of conformational landscapes. In this approach, a swarm of replica simulations interact cooperatively via a set of pairwise potentials incorporating attractive and repulsive components. Here, we apply the sesMD approach to explore conformations of the DFG motif in the protein, p38α mitogen-activated protein (MAP) kinase. In contrast to multiple MD simulations, sesMD trajectories sample a range of DFG conformations, some of which map onto existing crystal structures. Simulated structures intermediate between DFG-in and DFG-out conformations are predicted to have druggable pockets of interest for structure-based ligand design.
Introduction

Crystal structures are able to provide molecular level detail on the interaction of ligands with their target receptors. Recently, valuable additional information has been mined from computing the time evolution of an initial protein or protein-ligand crystal structure using molecular dynamics (MD) simulation. For example, MD simulations of HIV integrase discovered a previously unknown part of the binding site that was able to explain experimental binding and mutagenesis data and helped lead to the successful development of the drug, raltegravir, now a frontline drug in AIDS treatment.\(^1\) Similarly, potential design opportunities from MD have been found for example for neuraminidase\(^2\) and in recent work on \textit{T. cruzi} trans-sialidase.\(^3,4,5\)

However, due to their rough free energy surfaces, observing non-trivial changes in protein structure, for example the opening and closing of pockets, loops and grooves, is not commonly possible by MD without application of biasing potentials.\(^6\) These biasing methods can be equilibrium (e.g. umbrella sampling)\(^7\) or non-equilibrium (e.g. steered MD) approaches.\(^8\) Unfortunately, the manner in which these biasing potentials or forces should be applied is not always apparent. Specific biased MD approaches range in their complexity of application and computational overhead, and include active site pressurization,\(^9\) locally enhanced sampling,\(^10\) accelerated molecular dynamics\(^11\) and replica exchange-based approaches\(^12\) as well as the group formed by local elevation,\(^13\) conformational flooding\(^14\) and metadynamics methods.\(^15,16\)

Here, we propose an approach to exploring conformational space utilising a swarm-based MD scheme. This multicopy MD algorithm builds on the SWARM-MD method, first
The SWARM-MD approach involves creating multiple copies of the molecular system (the swarm) and then simulating the dynamics of each copy in parallel. The simulations of these copies mimic cooperative swarm behaviour through the addition of attractive forces that act on each member of the swarm driving their trajectories towards the mean trajectory of the entire swarm. In this way, the additional attractive force experienced by a given replica within the swarm is directly proportional to its distance from the average swarm structure. We demonstrated that our modified version of SWARM-MD is a simple and effective approach to optimization of peptide and protein structures. For example, during twenty independent 40 ns MD simulations of the Trp-cage miniprotein, none were able to fold to within a backbone RMSD of 1.5 Å of the NMR conformation. By contrast, of twenty SWARM-MD replicas, sixteen folded to a backbone RMSD below 1.5 Å within 40 ns.

In this work, we move from using a swarm of MD replicas to optimise biomolecular structure to instead explore the conformational landscape. To promote barrier crossing into alternative energy minima, we introduce a set of attractive and repulsive pair potentials which act between swarm copies. We refer to this method subsequently as swarm-enhanced sampling MD (sesMD).

We evaluate the ability of this sesMD approach to enhance conformational exploration of a protein known to exhibit different structures, the anticancer target p38α mitogen activated protein (MAP) kinase. Specifically, we focus on sampling of its Asp168-Phe169-Gly170 (DFG) motif, which can be characterized by different orientations of its phenylalanine sidechain: in the “DFG-out” conformation, Phe169 points away from the nearby αC helix and projects into the ATP binding pocket, exposing an additional hydrophobic cavity,
sometimes referred to as the allosteric or deep pocket. In the “DFG-in” conformation, access to the deep pocket is hindered by Phe169 (Figure 1). Whilst active and inactive states of protein kinases can exhibit DFG-in conformations, the DFG-out conformation is ordinarily associated only with the inactive state. Thus, the movement of the phenylalanine sidechain to the DFG-out conformation creates this deep pocket, a cavity which has been exploited in design of inhibitors with good specificity, i.e. by type II kinase inhibitors, such as imatinib.

Experimental and computational studies have found that there is a significant energetic barrier between kinase DFG-in and out states. For example, molecular dynamics simulations have required strategies such as studying kinase mutants with modified kinetics, massively distributed MD simulations or biasing methods.

Here, we apply sesMD to both DFG-out and DFG-in crystallographic conformations of p38α MAP kinase. We assess the ability of sesMD to explore alternative conformations of the kinase that could be targeted subsequently by structure-based drug design approaches.

Methods

sesMD. The SWARM-MD method described by Huber and van Gunsteren involves the attraction of swarm members to the mean of the swarm via an extra potential, \( V^{swarm} \), given by:

\[
V^{swarm}(\{\phi^\alpha\}) = \sum_{\alpha} A \exp\left[-Bd^{rms,av}_\alpha (\phi^\alpha)\right]
\]

(1)

where \( d^{rms,av}_\alpha \) is the root-mean-square dihedral angle difference of swarm member \( \alpha \) from the average location of the swarm, \( \phi^\alpha \) is set of dihedral angles of swarm member \( \alpha \) to which the
swarm potential is applied and $M$ is the number of members in the swarm. $A$ and $B$ are parameters that govern the strength and range of attraction between each member and the swarm average.

Here, we adapt SWARM-MD from an optimisation approach to a method that permits enhanced coverage of conformational space, such that the multiple MD simulations of a molecular system interact in a cooperative swarm-like fashion to surmount high energy barriers between conformations. We therefore introduce the following swarm-enhanced sampling (ses) potential, $V_{ses}$:

$$V_{ses} (\{\phi^\alpha\}) = \sum_{\alpha} \sum_{\beta \neq \alpha} \left( A \exp \left[ -B d_{rms}^{\alpha \beta} (\phi^\alpha, \phi^\beta) \right] + C \exp \left[ -D d_{rms}^{\alpha \beta} (\phi^\alpha, \phi^\beta) \right] \right)$$  \hspace{1cm} (2)

where $A$ – $D$ are suitably calibrated parameters for attractive (A,B) and repulsive (C,D) terms; and $d_{rms}^{\alpha \beta} (\phi^\alpha, \phi^\beta)$ is the root-mean-square dihedral angle distance between swarm members $\alpha$ and $\beta$. Distinct from eq. 1, the first term in eq. 2 is a pairwise attractive potential, acting between pairs of swarm replicas. This potential is balanced by a repulsive exponential potential (the second term in eq. 2). In this scheme, the values of parameters $A$ and $C$ must be negative and positive, respectively, while both $B$ and $D$ should be positive. The potential seeks to increase the spread of conformations explored, whilst promoting transitions between wells. The resulting dynamics are therefore a combination of (i) the ses potential, $V_{ses}$, applied to a selected set of dihedral angles and acting between pairs of replicas, and (ii) the interatomic potentials of the molecular mechanics force field, which act individually on each member of the swarm.
As a simple illustration of sesMD, we consider the effect of introducing $V_{ses}$ on sampling the conformational space of two solutes - pentane and alanine dipeptide. Firstly, eight unbiased MD simulations of 10 ns were performed of each solute in explicit aqueous solvent (for computational details, see Supporting Information). Sampling of low energy regions for pentane (Figure 2a) and alanine dipeptide (Figure 2c) remains fairly localised, with only minor excursions into alternative wells on the landscapes. Alternatively coupling together the eight replicas via the $ses$ potential enables exploration of alternative low energy regions during 10 ns of sesMD: for pentane, much greater coverage of symmetry-related gauche conformations is found (Figure 2b); for alanine dipeptide, sampling of other regions of the Ramachandran landscape is achieved (Figure 2d). However, we note that, due to the presence of the $V_{ses}$ term acting between replicas, the configurations sampled by sesMD are not drawn from a Boltzmann distribution. Nevertheless, the underlying potential energy surface topologies appear to be reasonably preserved via sesMD sampling (Figure 2b,d). Furthermore, due to the well-defined nature of the overall swarm Hamiltonian, it is formally possible to recover Boltzmann-weighted properties from the sesMD method where required (for more details, see Supporting Information).

**Computational details.** As starting points for molecular dynamics simulations of p38α MAP kinase, the crystallographic structures 1P38 and 1WBT were used, representing DFG-in and DFG-out conformations of the protein respectively. The structures were prepared for MD using the xleap program from the AMBER 11 suite. The systems were modelled using the AMBER ff99SB force field and solvated with ~13000 TIP3P water molecules. Counterions were added in order to neutralize the net charge of the systems. The SHAKE algorithm constrained bonds between hydrogen and heavy atoms. A 2 fs time step was used. Simulations were performed in the NPT ensemble with a Langevin thermostat, using a
collision frequency of 2 ps$^{-1}$. Isotropic position scaling was used to maintain an average pressure of 1 atm. The particle mesh Ewald method$^{38}$ was used for long range electrostatic interactions and a 12 Å cut-off for long-range non-electrostatic interactions.

For sesMD, the $ses$ potential described in eq. 2 was applied to the $\phi$ and $\psi$ backbone dihedral angles of the DFG motif (Asp168, Phe169, and Gly170). For sesMD simulations starting from the DFG-in 1P38 crystal structure, the sesMD parameters (A,B) and (C,D) were taken as (-150.0 kcal/mol, 0.2 rad$^{-1}$) and (150.0 kcal/mol, 0.8 rad$^{-1}$) respectively. For sesMD simulations initiated from the DFG-out 1WBT structure, the parameters (A,B) and (C,D) adopted were (-250.0 kcal/mol, 0.2 rad$^{-1}$) and (375.0 kcal/mol, 0.8 rad$^{-1}$) respectively. The two $ses$ potentials therefore differ in their dependence on $d_{\alpha\beta}^{rms}$ (Figure 1S). The choice of parameters for the two structures is to some degree a reflection of the local potential energy landscape of the DFG loop within the two kinase states of DFG-in and DFG-out: a stronger $ses$ potential was able to be applied to the DFG-out structure whereas, apparently due to the greater steric constraints of the occupied deep pocket within the DFG-in protein structure, the same strength of potential led to unphysical deformation of peptide bonds within the DFG loop.

For sesMD, 24 replicas of the kinase were employed. Initial configurations of the 24 replicas were generated as follows: the final structure of a 700 ps MD simulation at 300 K was taken as the initial configuration for 24 independent replica 500 ps simulations at 300 K with different initial velocities. SesMD simulations were then performed using these 24 configurations: during the first 500 ps of sesMD, the parameters A-D (eq. 2) were increased linearly from zero to their final values, followed by 500 ps of sesMD at the final A-D values; these parameters were then reduced to zero over 500 ps, and followed by 3.5 ns of unbiased
MD. The 24 independent 5 ns MD simulations used for comparison to the sesMD simulations were initiated from the configurations immediately prior to first application of the ses potential detailed above. All MD and sesMD simulations were performed using a modified version of sander from the Amber 11 molecular simulation package. Configurations were archived every 2 ps for analysis.

For principal component analysis (PCA), conformations were taken from every tenth frame of each sesMD simulation replica’s trajectory and fitted onto the Cα atoms of an average conformation calculated across all replicas in that simulation. Using these fitted conformations, PCA of the atomic displacements of the DFG residues was performed using the ptraj module of Amber 11. Clustering analysis of the DFG residues was performed using the kclust algorithm from the MMTSB toolset, with a fixed clustering radius of 4.0 Å. Pocket detection and druggability analysis were performed using PocketAnalyzer and Site Finder from MOE.

**Results**

We evaluate the ability of sesMD to sample the conformational plasticity of apo p38α MAP kinase, initiated from either a DFG-out crystal structure (1WBT) or DFG-in crystal structure (1P38). As described in Methods, we apply the ses potential of eq. 2 to the six backbone torsions linking the DFG loop amino acids, performing 5 ns of sesMD simulation using 24 replicas of the kinase in explicit solvent. These 5 ns of simulation comprise 1.5 ns under the influence of the ses potential, followed by 3.5 ns of unbiased MD. In the following discussion, we compare these sesMD trajectories with 24 independent 5 ns MD simulations of the kinase in explicit solvent.
Sampling from DFG-out and DFG-in crystal structures. We consider the ability of MD and sesMD to sample the backbone φ and ψ torsions for the three residues of the key DFG motif of the kinase. As a reference, based on the annotation of the MOE kinase explorer database, we plot the φψ angles of the 74 DFG-in p38α MAP kinase crystal structures (green, Figure 3), 61 DFG-out structures (red, Figure 3) and the 33 structures denoted as “unclassified” (blue, Figure 3). The DFG-in and DFG-out distributions of crystallographic φψ angles are largely distinct for all three residues of the loop, but for Asp168 in particular (Figure 3a). Thus, the φψ angles of Asp168 in the DFG-out 1WBT structure are (-139°, 111°), which sit within the DFG-out cluster around the β-sheet region (Figure 3a); whereas, for 1P38, the φψ angles are (32°, 68°) lying within the narrow left-handed α-helical region of DFG-in structures of the map (Figure 3a).

Unbiased MD simulations initiated from the 1WBT structure appear only to explore φψ values in the locality of its initial backbone conformation (Figure 3d). However, application of the ses potential leads additionally to exploration of the DFG-out φψ cluster in the right-handed α-helical region, and, notably, around the DFG-in cluster as well (Figure 3a). Broader sampling by sesMD over MD is also observed for the Phe169 (Figure 3b versus Figure 3e) and Gly170 backbone torsions (Figures 3c versus Figure 3f); generally, the crystal φψ values map on to the conformations sampled by sesMD although some simulated conformations do encompass regions of the Ramachandran map with few crystal structures (eg. φ of 90° for Phe169; φ of 180° for Gly170).

Interestingly, for MD simulations initiated from the 1P38 DFG-in structure, φψ values of Asp168 for both DFG-in and DFG-out regions are explored (Figure 3j). However, once again, sesMD provides more complete coverage of φψ space (Figure 3g,h,i), such that, for
example, DFG-in $\varphi_\psi$ values in the region of (-90°,-45°) are explored for Asp168 and Phe168 (Figures 3g,h); and in the vicinity of (90°,0°) for Gly169 (Figure 3i).

To provide a more complete view of sampling of the DFG loop conformation, principal component analyses (PCA) were performed for all DFG heavy atoms, based on the 168 X-ray crystal structures in combination the 1WBT trajectory (Figure 4a) or 1P38 trajectory (Figure 4b) from the MD or sesMD methods. The two principal components from these PCAs account for 75% and 76% of the variance in atomic displacements of the DFG motif in the 1WBT and 1P38 simulations respectively. Projection of the 168 X-ray crystal structures onto the axes defined by these first two principal components (Figure 4a,b) results in two distinct clusters, separating DFG-out (red, Figures 4a,b) and DFG-in (green, Figure 4a,b) conformations. Conformations unclassified by the MOE kinase explorer database to some degree occupy both clusters (blue, Figure 4a,b).

It can also been seen that unbiased molecular dynamics simulations explore only locally around the initial structures of the trajectories, for either 1WBT (yellow, Figure 4a) or 1P38 simulations (yellow, Figure 4b). By contrast, sesMD simulations of 1WBT and 1P38 explore more extensively, in both cases populating regions in the principal component space around the initial DFG geometry and the opposing DFG orientation (cyan, Figure 4a,b).

Further analysis of the sesMD trajectories indicates that the PC space most remote to the initial structures are sampled by a minority of its 24 replicas: this is shown by analysis of the root mean square deviation (RMSD) in atomic position of the DFG loop heavy atoms of each of the 24 replicas with respect to 1WBT and 1P38 X-ray structures when aligned according to the overall protein conformation (Figures 2S and 3S). For the majority of the replicas in the
1WBT DFG-out simulation, the RMSD values are ~2-3 Å from this initial structure (Figure 2S), although for two replicas which we label 1\textsubscript{out} and 2\textsubscript{out}, the RMSD relative to 1WBT rises to ~7 Å (Figure 5a and b respectively). Similarly, for sesMD simulations from 1P38, the majority of conformations sample locally (Figure 3S); however, for one replica of the 1P38 simulation, which we label 1\textsubscript{in}, the RMSD rises to ~7 Å from 1P38 (Figure 5d). If we chart the conformations sampled in the defined PC space by replicas 1\textsubscript{out}, 2\textsubscript{out} and 1\textsubscript{in}, we observe that they occupy the majority of remotely sampled regions in the PCA (orange and cyan, Figure 5c; magenta, Figure 5e).

For all three of these simulation replicas, the largest change in RMSD occurs during the application of the \textit{ses} potential rather than the free MD phase of the simulations. More specifically, it occurs during the phase in which there is gradual increase of the \textit{ses} potential, i.e. during the interval 0 to 0.5 ns (Figure 5a,b,d). We may also examine the RMSD of the DFG loops of these replicas with respect to their opposing crystal structure conformations. For replica 1\textsubscript{out}, the RMSD remains high with respect to the 1P38 structure throughout the simulation, dropping only slightly from 8 to 6 Å (green, Figure 5a). A similar pattern is observed for replica 1\textsubscript{in} of the 1P38 sesMD simulation (Figure 5d). Indeed, the high RMSD of replicas 1\textsubscript{out} and 1\textsubscript{in} with respect to the DFG-in and DFG-out crystal structures is suggestive of adopting an intermediate DFG geometry. However, for replica 2\textsubscript{out}, the RMSD drops over the course of the simulation to a value of ~3-4 Å (Figure 5b).

In order to obtain greater molecular detail of conformational change in these replicas, we consider two inter-residue distances, which have been suggested to be useful diagnostic measures of DFG-in and out conformations.\textsuperscript{32} These are defined as the distances between the side-chain centroids of Asp168 and Lys53 (neighbouring the glycine rich loop); and between
the sidechain centroids of Phe169 and Leu74 (located in the αC-helix). For the 1WBT DFG-out crystal structure, these distances are 13.5 and 17.8 Å respectively; and for the 1P38 DFG-in structure, the values are 11.8 and 6.9 Å respectively. Thus, from DFG-out to DFG-in, the considerable shortening of the Phe169-Leu74 distance and lesser lengthening of the Asp168-Lys53 distance indicates that Phe169 now occupies the kinase deep pocket, and Asp168 has correspondingly swung into the active site. For the first 2 ns of sesMD of 1WBT replicas 1\textsubscript{out} and 2\textsubscript{out}, we observe deviation of the Asp168-Lys53 distance from its crystallographic value transiently (Figure 6a), and a drop in Phe169-Leu74 distance to ~11 Å (Figure 6b). However whereas replica 1\textsubscript{out} remains at a Phe169-Leu74 distance of ~11 Å, replica 2\textsubscript{out} further decreases to a final value of ~7 Å, the crystallographic value of the DFG-in structure (Figure 6b). For replica 1\textsubscript{in} of 1P38, local conformational sampling is observed around the Asp168-Lys53 X-ray distances (Figure 6c). By contrast, a large change in Phe169-Leu74 distance is observed to stabilise in a proximity to the DFG-out X-ray value (Figure 6d).

We can visually trace the conformational motion of the DFG loop in the three replicas by considering snapshots from their trajectories. The DFG-out conformation of the 1WBT crystal structure (red, Figure 7a,b,c) and DFG-in pose of the 1P38 crystal structure (green, Figure 7a,b,c) bracket conformations sampled over the sesMD simulations. These sesMD snapshots show the time progression over the 5 ns trajectory (from orange to yellow, cyan, then blue, Figure 7a,b,c). For both replicas 1\textsubscript{out} and 2\textsubscript{out} (respectively Figure 7a,b), their trajectories show movement from their initial DFG-out structures (red) towards the opposing DFG-in conformation (green). However, whereas replica 1\textsubscript{out} arrests at an intermediate structure (Figure 7a), replica 2\textsubscript{out} completes the transition to a DFG-in geometry, such that Phe169 occupies the deep pocket (Figure 7b) and maps onto DFG-in orientations found in the apo (1P38) and liganded (2HVC, 3LGC)\textsuperscript{43,44} DFG-in crystal structures (Figure 8). This
The structural transition is more completely illustrated by a movie of the trajectory (Supporting Information, Movie S1). The DFG-out to in transition contrasts with a control simulation we ran from the same initial conditions of replica 2_{out}, but using 50 ns of unbiased MD. Here, no transition is observed, but instead the kinase remains in a local DFG-out conformation (Movie S2).

For replica 1_{in}, a partial transition from DFG-in to DFG-out is observed, such that there is a significant shift in Phe169 conformation (Figure 7c). However, Asp168 remains in a local well around its initial structure, maintaining a polar interaction with the backbone of Asn152.

For replicas 1_{out}, 2_{out} and 1_{in}, the largest movement of the DFG loop is associated with the period where the ses potential is increasing to its full strength over the interval 0.5 to 1.0 ns (orange, Figure 7). However, also particularly important for the replica 2_{out} simulation is the final 3.5 ns of non-sesMD, during which the activation loop relaxes to a structure within the manifold of DFG-in loop conformations that are observed crystallographically (Figure 4S). A relaxation of the swarm potential was found to be similarly important in the final stages of folding of Trp cage miniprotein by SWARM-MD.\textsuperscript{18}

Conformations and druggability of intermediate structures of the DFG loop. It is clear that a spectrum of DFG loop conformations is sampled during sesMD of p38α MAP kinase (for example, Figure 7), some of which may present opportunities for molecular design. To this end, we perform conformational clustering of the combined trajectories of all 24 sesMD replicas, for 1WBT and 1P38 simulations separately. Adopting a fixed clustering radius of 4.0 Å, we identify eight clusters of distinct conformations explored by the 1WBT sesMD replicas (labelled C1_{out} – C8_{out}) and five clusters of 1P38 sesMD conformations (C1_{in} – C5_{in}). From the location of these structures in PC space, we classify the clusters into DFG-out-like
and DFG-in-like conformations, for 1WBT (Figure 8a) and 1P38 simulations (Figure 8b). As might be expected, the most populated clusters are localised around the parent X-ray conformation. Thus, for the 1WBT sesMD simulation, clusters $C_{1_{\text{out}}}$ – $C_{4_{\text{out}}}$, representing 85% of the total sesMD configurations (Table 1), occupy DFG-out-like conformations (left-hand side of Figures 9a and 10); however, the less populated $C_{5_{\text{out}}}$-$C_{8_{\text{out}}}$ clusters occupy a spectrum towards a DFG-in structure (right-hand side of Figures 9a and 10). For the 1P38 simulation, 80% of the conformations fall into a $C_{1_{\text{in}}}$ DFG-in-like conformation (left-handside of Figure 9b; righthand of Figure 10, Table 1). Clusters $C_{2_{\text{in}}}$ and $C_{3_{\text{in}}}$, representing ~17% of conformations sampled, are characterized by a displacement of the Phe169 side-chain towards the $\alpha$C-helix. Clusters $C_{4_{\text{in}}}$ and $C_{5_{\text{in}}}$ group a smaller fraction of the ensemble, belonging only to replica $1_{\text{in}}$, which adopted the intermediate state described previously (Figure 7c).

From inspection, we observe similarities in transition structures sampled by the 1WBT sesMD simulation and crystal structures of p38$\alpha$ MAP kinase with intermediate DFG loop conformations. This is achieved despite the fact that these crystal structures are complexed with ligand, whereas here we perform sesMD of the kinase in the absence of ligand. Good agreement is seen in the DFG conformations of 15 crystal structures of the kinase and that of replica $1_{\text{out}}$ (Figure 11a). For example, sesMD conformations of Asp168 and Phe169 residues in this replica superimpose to within an RMSD of 0.4 Å of p38$\alpha$ MAP kinase in cocrystal with pyrrolotriazine compounds (pdb codes 3BV2 and 3BV3, Figure 11b).

Furthermore, other p38$\alpha$ MAP kinase transitional structures from sesMD bear a resemblance to intermediates identified from a high temperature restrained MD study of the kinase. In that work, pseudo-DFG-in and pseudo-DFG-out transition structures of p38$\alpha$ MAP kinase...
were defined, both lying at an RMSD of ~6-8 Å from DFG-in and DFG-out X-ray structures. These high temperature conformations were characterised by a β-turn-like structure, with the sidechain of Leu171 occupying the location normally occupied by Phe169 in a DFG-in (denoted pseudo-DFG-in) or DFG-out (pseudo-DFG-out) conformation. From the 1P38 simulation, we obtain sampling of structures akin to a pseudo-DFG-out conformation from sesMD, such that the side chain of Leu171 replaces the position previously occupied by the aromatic ring of Phe169 (Figure 11c); however, only a partial pseudo-DFG-in structure is found either from the 1WBT or 1P38 simulation (Figure 11d).

In assessing the opportunities for structure-based design against p38α MAP kinase, it is useful to characterise the pockets formed by the clustered sesMD conformations. For reference, we first consider the average pocket shapes arising from the 74 DFG-in and 61 DFG-out crystal structures: the location of the additional deep pocket is clearly shown in X-ray DFG-out structures relative to DFG-in (Figure 10). However, on average the total pocket volumes in the DFG-out and DFG-in crystal structure sets are very similar, with values of 784 and 736 Å³ respectively (Table 1). This hides considerable variability; for example, the volumes of pockets in 1WBT and 1P38 are 1440 and 989 Å³ respectively (Table 1).

A range of pocket shapes (Figure 10) and volumes (Table 1) are exhibited by the conformational clusters of p38α MAP kinase from sesMD, i.e. structures C1_out – C8_out and C1_in – C5_in. The pockets in C1_out – C4_out display a dual lobed density, reflecting the volumes of the hinge and deep pockets, whereas structures C1_in – C3_in possess a single pocket in the hinge region (Figure 10). These pocket shapes are akin to those found in the parent X-ray structures (Figure 10). Interestingly, the C5_out kinase conformation has a pocket shape distinct to that of C1_out – C4_out and X-ray structures. This cluster conformation shows an orientation
of Phe169 comparable to that encountered on one of two transition pathways sampled via the
aforementioned high temperature MD study.\textsuperscript{27} Clusters C\textsubscript{6out} – C\textsubscript{8out} clearly show diminished
deep pockets and expanded cavities in the hinge region typical of the average
crystallographic DFG-in pocket (Figure 10). For 1P38 sesMD conformations C\textsubscript{4in} and C\textsubscript{5in},
there is an opening up of the deep pocket region that is comparable to the topology of the
DFG-out pocket (Figure 10). Indeed, cluster C\textsubscript{5in} has the largest average pocket volume of all
the 1P38-derived clusters identified and closest in volume to that of the 1WBT structure
(Table 1).

To quantify the potential of these pockets for binding small molecules, we perform an
assessment of their druggability. Specifically, for the kinase structures closest to the cluster
centroids, we compute a propensity for ligand binding (PLB) index\textsuperscript{46} for pockets defined by
Site Finder (we note the kinase pocket topologies from Site Finder are broadly comparable to
that obtained by Pocketanalyzer\textsuperscript{PCA}). The PLB index is based on the specific amino acid
composition of a pocket and has proven effective in predicting drug binding propensities for
known protein structures.\textsuperscript{46} Here, we first consider the PLB indices for 1WBT and 1P38 X-
ray structures, which have values of 3.6 and 2.7 respectively (Table 1). These large positive
PLB values indicate that both sites are predicted as druggable, as expected; the DFG out
1WBT structure has a larger index, in part a function of the larger pocket volume compared
to 1P38 (Table 1).

This trend is also reproduced by the PLB scores predicted for the cluster centroid structures:
sesMD structures based on 1WBT show a reduced PLB index by switching to a DFG-in
structure, such that the PLP index of 3.9 for C\textsubscript{1out} decreases to 2.7 for C\textsubscript{8out} (Table 1); the
opposite trend is observed for 1P38 sesMD structures progressing towards a DFG-out state
(compare C5<sub>in</sub> with C1<sub>in</sub>, Table 1). Interestingly, PLB indices of greater than 4 are obtained from intermediate structures on either of these pathways, such as C4<sub>in</sub>, C5<sub>in</sub> and C4<sub>out</sub> (Table 1). These structures could represent interesting starting points for structure-based inhibitor design.

**Discussion and Conclusions**

From the preceding analysis, it is evident that application of swarm-enhanced sampling MD to the DFG motif of p38α MAP kinase provides a broader sampling of DFG conformational states compared to unbiased MD simulations. From sesMD of an initially DFG-out kinase conformation, we observe sampling of a range of loop conformations which include DFG-in structures (Figure 8) and intermediate geometries (Figure 11a-c) resembling p38α MAP kinase conformations identified previously from crystallography<sup>45</sup> and computation.<sup>27</sup> From sesMD based on a DFG-in crystal structure, a range of DFG loop conformations are sampled, although to a lesser extent than for the simulation based on a DFG-out conformation. In particular, whilst a partial transition is observed for Phe169, Asp168 remains localised in an interaction with the backbone of Asn152. This hydrogen bond may point to the implication of the protonation state of Asp168 in conformational change, as suggested in work by Shan et al. on Abl kinase.<sup>28</sup> In both 1P38 and 1WBT sesMD simulations, we identify interesting intermediate pocket shapes, for example C4<sub>in</sub> and C4<sub>out</sub> pockets, that could be useful in identifying novel kinase inhibitors via virtual screening.

SesMD is a biased MD approach using multiple simulation replicas. Whilst the concept of coupling replicas within a multicopy MD framework to improve sampling is not new,<sup>6,10,47,48,49,50,51,52</sup> in sesMD we have implemented a novel form of potential between replicas, incorporating attractive and repulsive terms. This swarm of simulation replicas could be
viewed as somewhat akin to a swarm of van der Waals particles, simultaneously exploring mutually exclusive regions of phase space. An alternative multicopy MD approach to enhanced sampling is the replica exchange method (REM),\textsuperscript{12} where at periodic intervals neighbouring replicas may be swapped between simulations. The key choice is in the proximity of the replicas in terms of temperature (for temperature REM) or potential (for Hamiltonian REM) such that efficient exchange can occur in accord with the requirement of detailed balance without requiring too many replicas to span the desired range of temperature or potential. This has proven to be an issue for large explicitly solvated solutes, which require a large number of replicas to achieve this overlap. This problem can be mitigated by using a new implementation of the replica exchange with solute tempering (REST2) method,\textsuperscript{53} where the replica Hamiltonians are scaled in order to remove the dependence on the number of explicit water molecules. SesMD is not subject to the energy overlap requirement of REM. However, in its current implementation, sesMD exchanges coordinate/dihedral information between all replicas at each timestep, in contrast to REM, which compares only neighbouring replicas typically at every ps. Due to this communication in sesMD, for a given timestep, the decrease in speed of sesMD compared to the equivalent set of independent MD trajectories is \textasciitilde38\%. However, for the 5 ns mixed sesMD/MD protocol applied here to p38\(\alpha\) MAP kinase, a decrease in speed of only 14\% was found, whilst the degree of sampling within the computed 5 ns was markedly improved for sesMD relative to 20 independent MD trajectories of 5 ns, or a single 50 ns trajectory.

From REM methods, as with the popular single-trajectory enhanced sampling approaches of accelerated MD\textsuperscript{11} and metadynamics\textsuperscript{15,16}, it is possible to obtain Boltzmann-weighted ensemble properties, directly or indirectly. This is also formally the case for sesMD: due to the well-defined Hamiltonian of the overall swarm of trajectories, one can reweight the swarm of
replicas according to the approach of Torrie and Valleau\textsuperscript{7} (see Supporting Information). We do not apply this approach here, where we scale the ses potential in seeking to drive conformational exploration of the kinase. Nevertheless, the general topology of the underlying unbiased potential energy surface appears to be preserved for sesMD of the model butane and dialanine systems (Figure 2b,d) and experimentally observed intermediate DFG conformations of the kinase are obtained during sesMD (Figure 11a,b).

We note that other biased MD approaches have been employed to study kinase conformational change: for example, targeted MD simulations have been used to study the structural reorganisation of Abl kinase\textsuperscript{54} and EGFR kinase.\textsuperscript{31} However, distinct from sesMD, a priori knowledge of the alternative protein conformation is required for targeted MD. Another biased MD study performed non-equilibrium MD simulations of p38\textsubscript{α} MAP kinase in order to explore DFG conformation.\textsuperscript{26} Steering forces were applied to the Asp168 and Phe169 side-chains of the DFG loop; the direction and magnitude of the pulling forces required recalibration at different stages in the conformational change. As with this non-equilibrium MD approach and the metadynamics method, for sesMD, a choice must be made as to the degrees of freedom to which to apply the biasing potential; although, in the case of the kinase, this is intuitive, \textit{i.e.} DFG backbone torsions. Potentially, for a given protein crystal structure, the choice of the amino acid dihedral set to explore by sesMD could be guided by experimental insight, for example from active site B-factors, or systematically mapped for active site residues. Accordingly, it will be important to assess the level of transferability of ses parameters to other systems of interest, for small and larger molecules. We note that careful system-dependent parameterisation is required for other biasing methods such as metadynamics and accelerated molecular dynamics. For the latter, an initial estimate of the energy landscape to explore is obtained from trial simulations in the absence of the biasing potential, prior to iter-
ative short biased simulations to obtain suitable parameters. We anticipate that future work will provide insight into the required parameter ranges and protocols for optimal sesMD applications.

SesMD therefore constitutes a straightforward, parallel and promising approach to conformational sampling. Here, we have applied sesMD to generating a range of possible alternative conformations of a protein based on a single crystal structure. Such a strategy applied to a therapeutic target protein could reveal novel protein concavities, of potential use in providing new directions for structure-based inhibitor design.

Acknowledgements

We acknowledge the support of the N8 HPC facility, BBSRC project BB/H016139/1 and Marie Curie project 275063.

Supporting information

This includes computational details for sesMD simulations of pentane and dialanine; a scheme for reweighting properties obtained from sesMD simulations; and further information on sesMD of p38α MAP kinase (biasing potentials, trajectory of replicas, activation loop conformation). This material is available free of charge via the Internet at http://pubs.acs.org.

References


(42) Chemical Computing Group MOE. *Montreal Quebec Canada 2003*.


Table 1. Population (pop) and average volume ($V_{av}$) of the binding pockets identified by PocketAnalyzer$^{PCA}$ for each cluster calculated in the clustering analysis along with an approximate classification of its DFG conformation (DFG conf). Standard deviations in parentheses. Propensity for ligand binding (PLB) of structures possessing the lowest distance from each cluster centroid, calculated by Site Finder.

<table>
<thead>
<tr>
<th>cluster</th>
<th>DFG conf</th>
<th>pop (%)</th>
<th>$V_{av}$ ($\text{Å}^3$)</th>
<th>PLB</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1$_{out}$</td>
<td>out</td>
<td>25.2</td>
<td>1212 (347)</td>
<td>3.9</td>
</tr>
<tr>
<td>C2$_{out}$</td>
<td>out</td>
<td>23.9</td>
<td>1044 (255)</td>
<td>3.0</td>
</tr>
<tr>
<td>C3$_{out}$</td>
<td>out</td>
<td>18.8</td>
<td>1334 (294)</td>
<td>4.3</td>
</tr>
<tr>
<td>C4$_{out}$</td>
<td>out</td>
<td>17.6</td>
<td>1255 (225)</td>
<td>4.5</td>
</tr>
<tr>
<td>C5$_{out}$</td>
<td>intermed</td>
<td>6.6</td>
<td>866 (278)</td>
<td>3.8</td>
</tr>
<tr>
<td>C6$_{out}$</td>
<td>intermed</td>
<td>3.9</td>
<td>712 (277)</td>
<td>3.5</td>
</tr>
<tr>
<td>C7$_{out}$</td>
<td>in</td>
<td>1.9</td>
<td>1181 (294)</td>
<td>2.8</td>
</tr>
<tr>
<td>C8$_{out}$</td>
<td>in</td>
<td>2.0</td>
<td>1071 (291)</td>
<td>2.7</td>
</tr>
<tr>
<td>C1$_{in}$</td>
<td>in</td>
<td>79.6</td>
<td>880 (260)</td>
<td>3.0</td>
</tr>
<tr>
<td>C2$_{in}$</td>
<td>in</td>
<td>14.0</td>
<td>686 (211)</td>
<td>2.9</td>
</tr>
<tr>
<td>C3$_{in}$</td>
<td>in</td>
<td>3.5</td>
<td>955 (242)</td>
<td>4.0</td>
</tr>
<tr>
<td>C4$_{in}$</td>
<td>intermed</td>
<td>2.6</td>
<td>770 (183)</td>
<td>4.5</td>
</tr>
<tr>
<td>C5$_{in}$</td>
<td>intermed</td>
<td>0.4</td>
<td>1104 (213)</td>
<td>4.4</td>
</tr>
<tr>
<td>DFG-in X-ray av</td>
<td>in</td>
<td>-</td>
<td>736 (246)</td>
<td>-</td>
</tr>
<tr>
<td>DFG-out X-ray av</td>
<td>out</td>
<td>-</td>
<td>784 (316)</td>
<td>-</td>
</tr>
<tr>
<td>1P38</td>
<td>in</td>
<td>-</td>
<td>989</td>
<td>2.7</td>
</tr>
<tr>
<td>1WBT</td>
<td>out</td>
<td>-</td>
<td>1440</td>
<td>3.6</td>
</tr>
</tbody>
</table>
List of figures

Figure 1. Superposition of 1P38 (green) and 1WBT (red) X-ray crystal structures of p38α MAP kinase. Backbone atoms are represented in cartoon while residues Asp168 and Phe169 are in sticks.

Figure 2. Distribution of pentane conformations in water (relative energies, in kcal/mol), defined by central C-C dihedral angles \( \phi_1 \) and \( \phi_2 \), during (a) eight independent MD simulations of 10 ns and (b) an eight-replica sesMD simulation of 10 ns. Distribution of dialanine conformations in water as a function of backbone dihedral angles \( \phi \psi \) during (c) eight independent MD simulations of 10 ns and (d) an eight-replica sesMD simulation of 10 ns.

Figure 3. Ramachandran plots of \( \phi \psi \) angles for each DFG residue, sampled by 24 replicas of sesMD simulations (black) starting from p38α MAP kinase crystal structures 1WBT (a,b,c) and 1P38 (g,h,i); and sampled by 24 independent MD simulations (d,e,f,j,k,l). Also shown are \( \phi \psi \) angles of 61 DFG-out (red), 74 DFG-in (green) and 33 unclassified (blue) p38α MAP kinase X-ray structures as defined in MOE kinase explorer database.

Figure 4. Projections of MD (yellow) and sesMD (cyan) ensembles of kinase and DFG-out (red), DFG-in (green) and unclassified (dark blue) p38α MAP kinase crystal structures onto space defined by top two principal components based on simulations starting from (a) 1WBT and (b) 1P38 crystal structures.

Figure 5. Time series of heavy atom RMSD (in Å) of DFG motif from sesMD simulations of p38α MAP kinase for replicas (a) l_out, (b) 2_out and (d) l_in, calculated against DFG-out
1WBT (red) and DFG-in 1P38 (green) crystal structures, during pre-sesMD equilibration (-0.5 to 0.0 ns) and sesMD trajectories (0.0 to 5.0 ns). Projections onto first two principal components of sesMD configurations of replicas (c) 1_out (orange) and 2_out (cyan); and (e) 1_in (magenta); alongside PC values for 24 MD replicas (yellow), DFG-in (green), DFG-out (red) and unclassified (dark blue) crystal structures.

**Figure 6.** Time series of distance between side-chain centroids for Asp168-Lys53 of (a) replicas 1_out (orange), 2_out (cyan) and (c) 1_in (magenta) of p38α MAP kinase, from pre-sesMD equilibration (-0.5 to 0.0 ns) and sesMD trajectories (0.0 to 5.0 ns). Time series of distance between side-chain centroids for Phe169-Leu74 of (b) replicas 1_out (orange), 2_out (cyan) and (d) 1_in (magenta) from pre-sesMD equilibration and sesMD trajectories. X-ray distances for 1WBT (red) and 1P38 (green) structures also shown for reference.

**Figure 7.** Conformation of Phe169 in selected snapshots from (a) replica 1_out and (b) replica 2_out from sesMD simulations of p38α MAP kinase. Conformations from sesMD are colour coded according to simulation time/stage: 0.5 – 1 ns (orange), 1 – 1.5 ns (yellow), 1.5 – 2 ns (cyan), 2 – 5 ns (blue). Also shown are DFG-in (1P38, green) and DFG-out (1WBT, red) crystal structures.

**Figure 8.** Final conformation from sesMD of replica 2_out showing Asp168 and Phe169 sidechains in p38α MAP kinase active site (blue), compared with its initial pose from DFG-out 1WBT X-ray structure (red); DFG-in crystal structures 1P38 (green), 2LGC and 3HVC (both light grey) also shown.
Figure 9. Projections onto space defined by first two principal components of ensemble of 24 sesMD simulation replicas, initiated from (a) 1WBT and (b) 1P38 crystal structures. Projections are colour coded according to cluster: C1_{out} and C1_{in} (red), C2_{out} and C2_{in} (orange), C3_{out} and C3_{in} (yellow), C4_{out} and C4_{in} (green), C5_{out} and C5_{in} (cyan), C6_{out} (blue), C7_{out} (magenta) and C8_{out} (purple).

Figure 10. Average binding pocket topologies found in structures within conformational clusters identified from sesMD simulations of p38α MAP kinase (using PocketAnalyzer^PCA), mapped onto structures nearest to cluster centroid. Each structure is colour coded according to cluster (see caption of Figure 8). The pocket is colour coded according to the weighted average (high frequency - blue; low frequency - red). As reference, sets of DFG-in and DFG-out crystallographic structures were used.

Figure 11. Superposition of a selected p38α MAP kinase conformation from sesMD replica 1_{out} (blue) with (a) 14 p38α MAP kinase X-ray structures (purple, PDB codes in Table 1S) and (b) with X-ray structures 3BV2 and 3BV3 (purple). Superpositions of selected intermediate DFG conformations (blue) from sesMD replicas (c) 1_{out} and (d) 1_{in}. For reference, DFG-in (1P38, green) and DFG-out (1WBT, red) X-ray structures also shown.
TOC graphic