Available online at www.sciencedirect.com
ScienceDirect
Journal of Molecular Graphics and Modelling 26 (2007) 187–197
Journal of Molecular Graphics and Modelling
www.elsevier.com/locate/JMGM

A new methodology for the simulation of flexible protein–ligand interactions

James Garner [a], John Deadman [c], David Rhodes [c], Renate Griffith [b,*], Paul A. Keller [a,**]

[a] Department of Chemistry, University of Wollongong, Wollongong, NSW 2522, Australia
[b] School of Environmental and Life Sciences, University of Newcastle, Callaghan, NSW 2308, Australia
[c] Avexa Ltd., 576 Swan Street, Richmond, Vic. 3121, Australia

Received 7 July 2006; received in revised form 16 October 2006; accepted 15 November 2006
Available online 21 November 2006

Abstract

A methodology has been developed for the simulation of induced fit between a ligand and its target protein. It utilizes constrained molecular dynamics where atoms determined to be immobile from difference distance matrix studies are fixed. Application of this methodology to HIV-1 reverse transcriptase (RT) as the example target protein has demonstrated its robustness. Short simulation times are sufficient to achieve good refinement of docking poses resulting from exchange of structurally dissimilar inhibitors between crystal structures.

Keywords: Drug design; Molecular dynamics; Difference distance matrices; Flexible docking; Flexible protein models

1. Introduction

The use of computer aided molecular design for the development of better therapeutics and for the greater understanding of protein structure and function is well established and these techniques have already had a direct impact on the establishment of new therapeutics on the market. However, most such studies have used the modified 'lock-and-key' approach where the conformational diversity of the small molecule ligand is taken into account, but the 'high molecular weight' ligand binding site (e.g. enzyme, receptor) is considered rigid. While significant successes have arisen from such a strategy, it limits the outcomes available as it ignores a common situation—most enzymes and receptors are flexible and contribute to ligand binding with a significant conformation change (induced fit). While earlier studies might have been hampered by a lack of computational power to investigate the flexibility of large molecules, the situation is rapidly changing and consequently, new methodologies are beginning to emerge. The significant further development of such techniques is not only critical for the understanding of protein function and the development of novel ligands as potential medicinal agents, it is also required to complement the enormously increasing bank of data arising from structural genomic studies. At this stage, this field of protein flexibility in computer-aided molecular design is still in its infancy.

Conformational change upon ligand binding is fundamental in protein behaviour for both enzymes and receptors with knowledge of the mechanistic machinery of a protein permitting a greater understanding of its function. Important to the understanding of structure-function relationships are the large scale collective movements of proteins upon binding of substrates and/or inhibitors, agonists and/or antagonists [1]. The understanding of these complex processes and functions is currently only partially addressed. It is imperative that such dynamic actions are investigated as the ability to predict new conformations/pocket shapes for ligand binding allows better potential medicinal agents to be designed and investigated and is likely to lead to better activity through new binding interactions.

Therefore, the ability to use computer-aided drug design (CADD) techniques to predict conformational changes in macromolecules, e.g. proteins, opens enormous possibilities that can be extended to investigations into protein motion and mechanics, modes of action and blocking, and proposed

* Corresponding author.
** Corresponding author. Tel.: +61 2 4221 4692; fax: +61 2 4221 4287.
E-mail addresses: Renate.Griffith@newcastle.edu.au (R. Griffith), keller@uow.edu.au, paul_keller@uow.edu.au (P.A. Keller).

1093-3263/$ – see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2006.11.004

binding cavity conformations for both substrates/agonists and potential inhibitors/antagonists. The ability to successfully generate such a model leads to a superior understanding of the fundamental components of the protein 'machine' as well as opening up large tracts of ligand diversity for potential therapeutic intervention by enzyme inhibition or interference with receptor signalling.

The major reason that the application of CADD to the flexibility of proteins has been generally avoided is due to the enormous conformational space available to such systems resulting from the increasing number of degrees of freedom. Current methods apply constraints to restrict the range of flexibility investigated to a manageable level and these have been recently reviewed [2–4]. They include the development of structure-based pharmacophores using several protein structures to account for the flexibility [5]. Algorithm development for the dynamic rearrangement of amino acid side chains [6] and the use of weighted masses molecular dynamics (WMMD) for the prediction of mobile domains within proteins [1]. Each of these methods has its own limitations e.g. some are limited to investigations of singular conformational changes rather than trajectory path simulations, or pocket flexibility investigations being limited to amino acid chain movements which do not consider the potential for the protein backbone to change conformation.

We report here our first results using a new technique for the incorporation of protein flexibility into structure-based computer-aided molecular design studies. Fundamental to our approach is the incorporation of information obtained with the aid of difference distance matrix comparisons about individual atoms of a protein structure with respect to how much they move upon substrate or ligand binding. The ability of these atoms to either not move or move by up to a defined distance is then built into molecular dynamic runs which allow atoms to move within those defined constraints.

2. General methodological principles

Our methodology aims to simulate the induced fit including potential conformational movements of both the ligand and the protein in a manner allowing structure based CADD on reasonable time scales. Crucial to this is the incorporation of restrictions on the motion of specific amino acid residues as well as the incorporation of information derived from all possible sources of the protein under study, e.g. crystal structures, NMR structures and computer-aided design models. Hence, the dynamic space that is sampled by the simulation will correlate more strongly to known motions, leading to a shorter simulation run required, and a higher probability of simulating the 'real' ligand–target interaction. Difference distance matrices have been utilized to analyse the range of movement that occurs when a ligand binds to the target protein.

This methodology is broadly applicable, with accuracy mainly dependant on the number of structures available of the protein to be investigated. We have used the human immunodeficiency virus (HIV-1) enzyme reverse transcriptase (RT) (Fig. 1) as the example in this work, due to its extensive flexibility, the variety of structural classes of known inhibitors (most common examples see Fig. 1), and the large pool of crystal structures available for analysis. The majority of inhibitors crystallized with RT belong to the non-nucleoside RT inhibitor class (NNRTI) which allosterically bind in a flexible hydrophobic pocket (the non-nucleoside inhibitor binding pocket, NNIBP) which does not exist in the native enzyme, but is created upon inhibitor binding. All available crystal structures were utilized that meet general, non-RT specific,

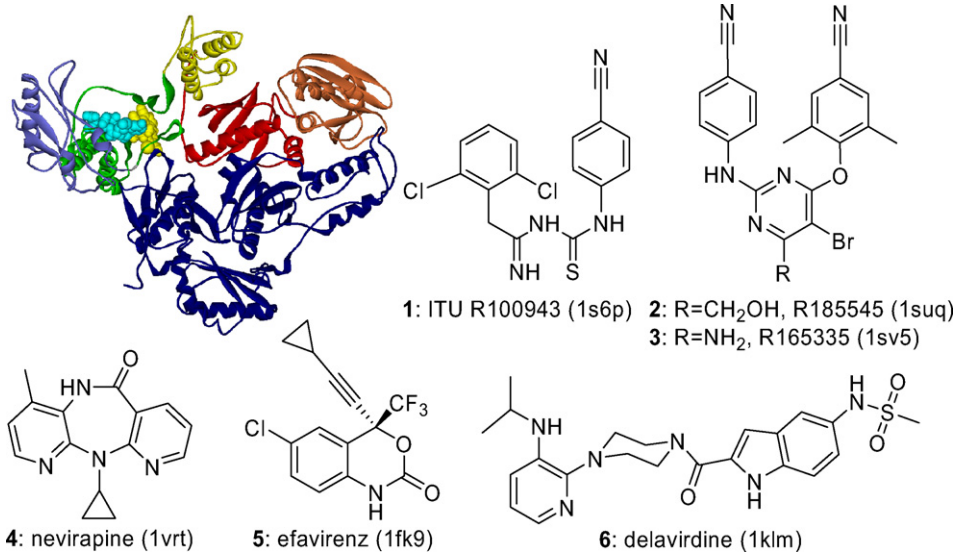


Fig. 1. The 'front' view of the structure of HIV-1 RT (ribbon representation). The commonly identified regions are indicated as follows; fingers (purple), palm (green), thumb (yellow), connection (red), RNase H (orange), and p51 subunit (dark blue). The location of a typical inhibitor (NNRTI)(yellow space filling) and the polymerase active site (PAS) (aqua space filling) are also shown (all atoms). Example NNRTIs investigated in this study are also shown. The source pdb files for these structures are indicated in parentheses.

requirements classified by resolution, completeness, and collection method.

2.1. Determination of immobile atoms

Using our implementation [7] of the well-documented difference distance matrix method [8], the movements of atoms/residues that occur after a ligand binds to a protein can be described *via* comparison to the unliganded protein. The method can also be used to group atoms that move in any defined range (e.g. 0–2 Å, 2–5 Å, etc.), but the primary function is to find those atoms that never move upon the binding of any ligand. In this method an immobile atom has been defined as one that only moves between 0 and 2 Å in all crystal structures. The immobile atoms are then fixed during the molecular dynamics simulation, cutting down the computational time required. Additionally, this subset of atoms is used to superimpose structures for analysis. This superimposition method has the benefit of being impartial, removing any bias occurring when arbitrarily defined superimposition subsets are used [9–12]. Whole protein simulations in conjunction with whole protein superimpositions for analysis of results can then be performed.

2.2. Progression of the methodology

The methodology was tested against systems of increasing complexity, in order to understand its limitations and to enable rapid modification of the methodology when problems were observed. Firstly, a selection of crystal structures were 'relaxed' investigating the stability of ligand/protein complex crystal structures when subjected to our simulations. Secondly, ligand exchanges were then performed attempting to model the induced fit binding mode. Initially, ligands were exchanged between crystal structures containing ligands from the same structural class, then between crystal structures containing structurally unrelated ligands. Thirdly, modelling the induced fit with mutant enzyme species was examined.

3. Results

3.1. Determination of immobile atoms

Our original subset of immobile atoms determined using difference distance matrices with our in-house software (Proflex 2.0) [7] was reassessed because of the larger pool of structures now available (67 versus 19). Stricter selection criteria for construction of this subset could be applied and consequently, only 10 of the 19 crystal structures originally examined were used in this study. The average resolution of the new crystal structure set is 2.77 Å (standard deviation (S.D.) 0.21), which is a modest improvement over the original set (2.85 Å, S.D. 0.35). The large size of the RT protein (nearly 1000 amino acids) does not allow comparison of all-atom distance matrices of all 67 crystal structures. Thus, we only compared C_α atoms. The subset of immobile residues is comprised of 302 C_α atoms, a decrease of 20% from the previous subset (376 C_α atoms), exhibiting 70% identity to the original subset, with most of the differences (66%) occurring in the p51 subunit. A cluster analysis of the subsets formed by comparing crystal structures with the same crystal packing (space group), showed that there were no significant differences in the subset caused by crystal packing artefacts. This refinement indicates that if only a few crystal structures are available, a robust subset of immobile residues can still be identified, as a large increase in the number of structures analysed has not resulted in a fundamentally different subset of immobile residues (here subsequently referred to as the superimposition subset).

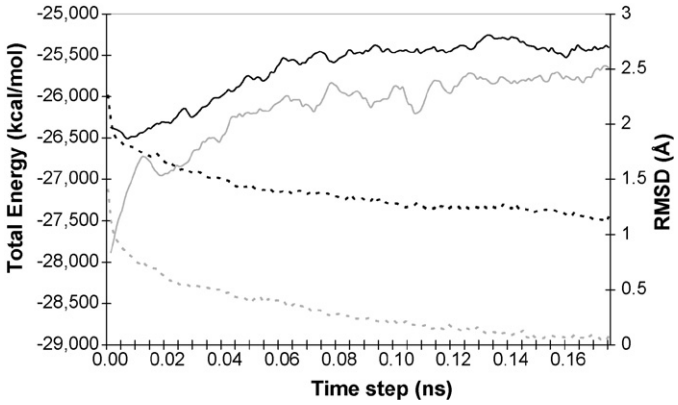


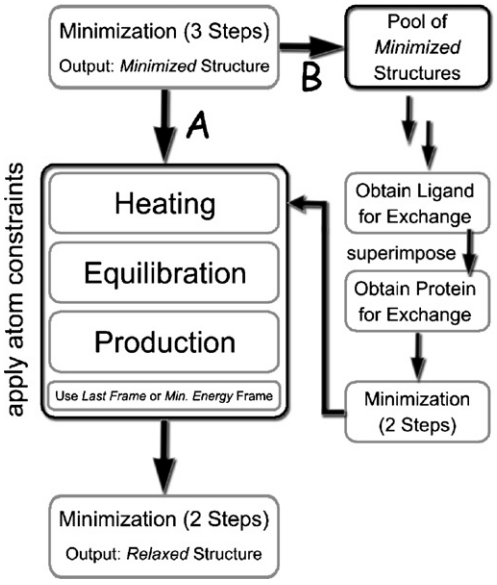
Fig. 2. Comparison between the simulations of the constrained (black) and unconstrained (gray) molecular dynamics (production) step of 1s6p. The RMSD is indicated by solid lines, the total energy by dashed lines.

3.2. Calibration of simulation time

A balance needed to be struck between computation time and completeness for the methodology development. The length of the molecular dynamics simulation was chosen such that the total energy and root mean square difference (RMSD)[1] had both appeared to reach equilibrium during the production step (an observed plateau over the last 0.025 ns). The structure chosen for the calibration of simulation time was 1s6p, which is crystallized with ligand **1** known to have high activity against native HIV-RT, but poor activity against strains with common mutations [13].

The optimum simulation time that met these criteria was determined to be 0.175 ns. Comparative runs were performed on an unconstrained system and a system with fixed atom constraints applied to the residues of the immobile subset as determined above. The constraints smoothed fluctuations in the RMSD and improved the time to convergence, even though the change in total energy took a similar path in both systems (Fig. 2). As a practical consideration, while the determination of the immobile residues was only performed on the C_α atoms (see above) the fixed atom constraint in the dynamics simulation was applied to the whole residue in the superimposition subset.

[1] The RMSD of the production is calculated by DS Modelling 1.1 by comparing each frame with the coordinates of the molecule just before the production stage begins. The comparison is done using all atoms and the conformations are superimposed first in order to calculate the RMSD with the maximum overlap of all atoms.



Scheme 1. Flow diagram summarizing the modelling protocols employed. The protocol for the relaxation of the crystal structures proceeded *via* path A, while the protocol for the ligand exchange simulations proceeded *via* path B.

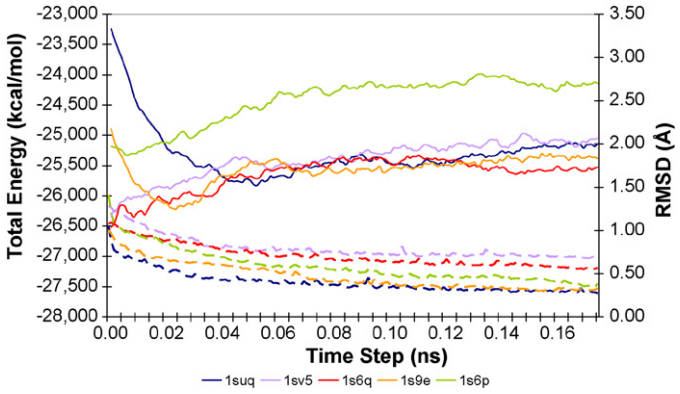


Fig. 3. Comparison of the total energy (left scale, dashed lines) and RMSD (right scale, solid lines) of the production step of the relaxation of five crystal structures.

3.3. Method development

3.3.1. Stability of crystal structures

Five RT crystal structures (1suq, 1sv5, 1s6q, 1s9e, 1s6p) [13] were analysed to investigate the extent of protein–ligand motion over the sampling time of the dynamics simulation. These crystal structures were selected due to their good resolution (average of 2.87 Å) and complete sequence of residues, allowing simulations to be performed on structures that did not need to be artificially completed. An eight-step experiment was employed (Scheme 1, path A), comprising a three-step minimization (with the output designated the *minimized* structure), followed by a three-step dynamics protocol and two-step minimization (with the output designated the *relaxed* structure). Fixed atom constraints were applied to the immobile atoms during the heating, equilibration, and production steps of the dynamics protocol. Each *relaxed* structure was compared by superimposition (using the C_α atoms of the superimposition subset) to the *minimized* structure. In each of the simulations the total energy and RMSD both reached an acceptable equilibrium during the production step (Fig. 3), with the higher observed RMSD of 1s6p attributed to it being crystallised with the ligand with the lowest affinity in the series.

3.4. Method development

3.4.1. Minimization of structures after molecular dynamics

In an early version of the methodology, the final two-step minimization was performed on the last frame of the production step (Scheme 1). After comparison by superimposition, two of the *relaxed* structures had poorer superimpositions compared to the other structures, with an average RMSD of all the superimpositions of 1.74 Å (S.D. 0.67). Examination of the production trajectory of the poorer simulations indicated that the frame with the minimum energy occurred much earlier in the simulation compared to those structures with better superimpositions. Therefore, the methodology was modified such that the minimum energy structure was isolated from the trajectory of the production step, and then minimized with the two-step protocol to obtain the *relaxed* structure.

After this modification the average RMSD of superimposition was 1.11 Å (S.D. 0.12) which compares remarkably well to the average RMSD of both the equilibrium (1.78 ± 0.65 Å) and production steps (1.91 ± 0.23 Å), especially when performing superimpositions utilizing large atom sets.

3.5. Method development

3.5.1. Analysis and validation

The average movement of the various structural domains was consistent (Table 1), with larger atom movements centralized in domains that are known to move significantly, e.g. the thumb (2.72 Å, S.D. 0.91) and fingers (1.36 Å, S.D. 0.38). There was no significant rearrangement of the side chain atoms (1.48 Å, S.D. 0.16) as indicated by their similarity to the average movement of the backbone atoms (1.20 Å, S.D. 0.16) and all atoms (1.48 Å, S.D. 0.15). The NNIBP was also stable (0.85 Å, S.D. 0.23).

The robustness of an experiment was also examined by plotting the distribution of the count of atoms versus the

Table 1
Comparison of average atom movement vs. structural feature

Region	Average atom movement (Å)	S.D.
All atoms	1.48	0.15
Superimposition subset atoms	0.85	0.09
Remainder of protein	1.79	0.20
NNIBP[a]	0.85	0.23
Fingers[a]	1.36	0.38
Thumb[a]	2.72	0.91
Backbone	1.20	0.16
Sidechain	1.48	0.16

[a] Regions as defined by Ren et al. [14].

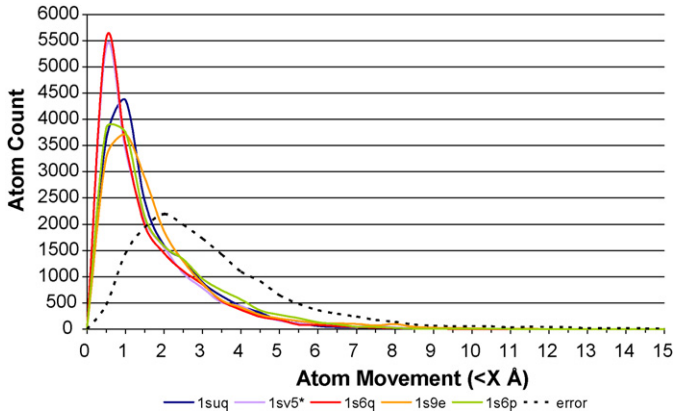


Fig. 4. Comparison of the atom count vs. atom movement after the relaxation of crystal structures. The atom movement distribution of 1s6q, which occurred after an atom typing error is indicated by the dashed line.

distances they moved (Fig. 4), e.g. 0–0.5 Å, 3.5–4 Å, during the relaxation simulation. This was performed as the RMSD of superimposition could still be acceptable even if there had been a significant problem in the dynamics simulation. A successful simulation generated a skewed distribution with most of the atoms moving <1 Å, while simulation problems could be identified by a significant deviation from this distribution. An example with 1s6q, occurred early in the methodology development, where despite an acceptable RMSD of 1.76, the atom movement distribution was so dissimilar to the other relaxations (Fig. 4, dashed line) that it prompted further investigation of the experimental conditions, exposing an initial atom typing error.

The scalar motion of every atom of the inhibitor binding pocket was calculated (Fig. 5) to compare changes in atom positions after the relaxation of the crystal structures. In general, the NNIBP was stable, with the ligand maintaining known π-stacking interactions with the important Tyr181 and Tyr188 residues. Movement of atoms <1 Å appeared to be random amongst all the structures studied.

There were three significant deviations observed within the binding pocket. Firstly, Leu100, Lys101, Val106 and Val179 had larger than average sidechain deviations attributed to the typical mobility of aliphatic sidechains. Secondly, the 3–5 Å movement of the sidechain of His235 is due to the reorientation of the protonated nitrogen of the imidazole.

Finally, after relaxation of the crystal structures a significant variation was observed in the position of Phe227, a residue that is also part of the β12–β13 hairpin (Phe227-His235), called the primer grip, and responsible for maintaining the primer terminus in the orientation required for nucleophilic attack on an incoming dNTP. The phenyl ring of Phe227 rotates from a 'down' to an 'up' position as exemplified with 1suq (Fig. 6). This was observed in many of the simulations performed during the methodology development, and was attributed to the simulation capturing a 'snapshot' of the dynamic equilibrium between two energetically favourable conformations. Interestingly, Phe227 exhibits different exposure in these conformations; in the 'down' position it is exposed from the 'back' of RT (Fig. 6, bottom left), and from the 'front' in the 'up' position (Fig. 6, top right). While the other residues of the NNIBP exhibit similar surface exposure when *minimized* and *relaxed*, Phe227 is hidden from the back of the *relaxed* structure (Fig. 6, bottom right). Additionally, the ligand is further exposed at the back of the pocket. This result may have implications on a possible mechanism for the entrance of the ligand into the binding pocket (which is presently unknown) as the rotation of this residue enlarges the solvent accessible portion of the NNIBP.

Overall the relaxation of the crystal structures indicated that only small differences occur in the protein conformation (and specifically the NNIBP) when compared to the *minimized* structures. These results are encouraging, suggesting that our molecular dynamics simulations, where around one third of the enzyme is fixed, do not distort the geometry of the original ligand-bound crystal structure.

3.6. Ligand exchange

3.6.1. Exchange of similar ligands

The ability of the methodology to use one crystal structure to predict the induced fit of different ligands was tested by choosing a pair of crystal structures from the relaxation study, exchanging the ligands, and applying the simulation protocol in an attempt to recreate the original situation. A modified protocol was used taking the minimized structures from the relaxation study for the exchange of a ligand, which was inserted by superimposition of the two crystal structures and deletion of unwanted elements (Scheme 1, path B). In a typical example the simulation of the induced fit from the exchange of the ligand (Fig. 1(**2**)) from 1suq into 1s6p had an RMSD of 1.70 Å when compared to *minimized* 1suq.

The ligand exchange analysis, using the same methods as the relaxation study, showed that with ligands of the same structural class, reproducible results on position and orientation of the ligand in the binding pocket could be achieved. There were only moderate changes in the position of the binding pocket residues, with major conformational changes isolated to the previously observed Phe227, Leu101 and Val179 movements.

3.7. Ligand exchange

3.7.1. Exchange of dissimilar ligands

A more practical application of the method was to attempt to simulate the induced fit of known ligands with available crystal structures, but from a different structural class. Nevirapine **4** and efavirenz **5** (Fig. 1) were selected as there are multiple high resolution structures available with these inhibitors bound. The crystal structures containing **4** (from 1fk9) and **5** (from 1vrt) were minimized using the three-step protocol and the ligands were superimposed into 1s6p after which the methodology (Scheme 1, path B) was applied to attempt to recreate the original induced fit (Fig. 7).

There was significant agreement between the original fit in the crystal structures and that predicted from the molecular

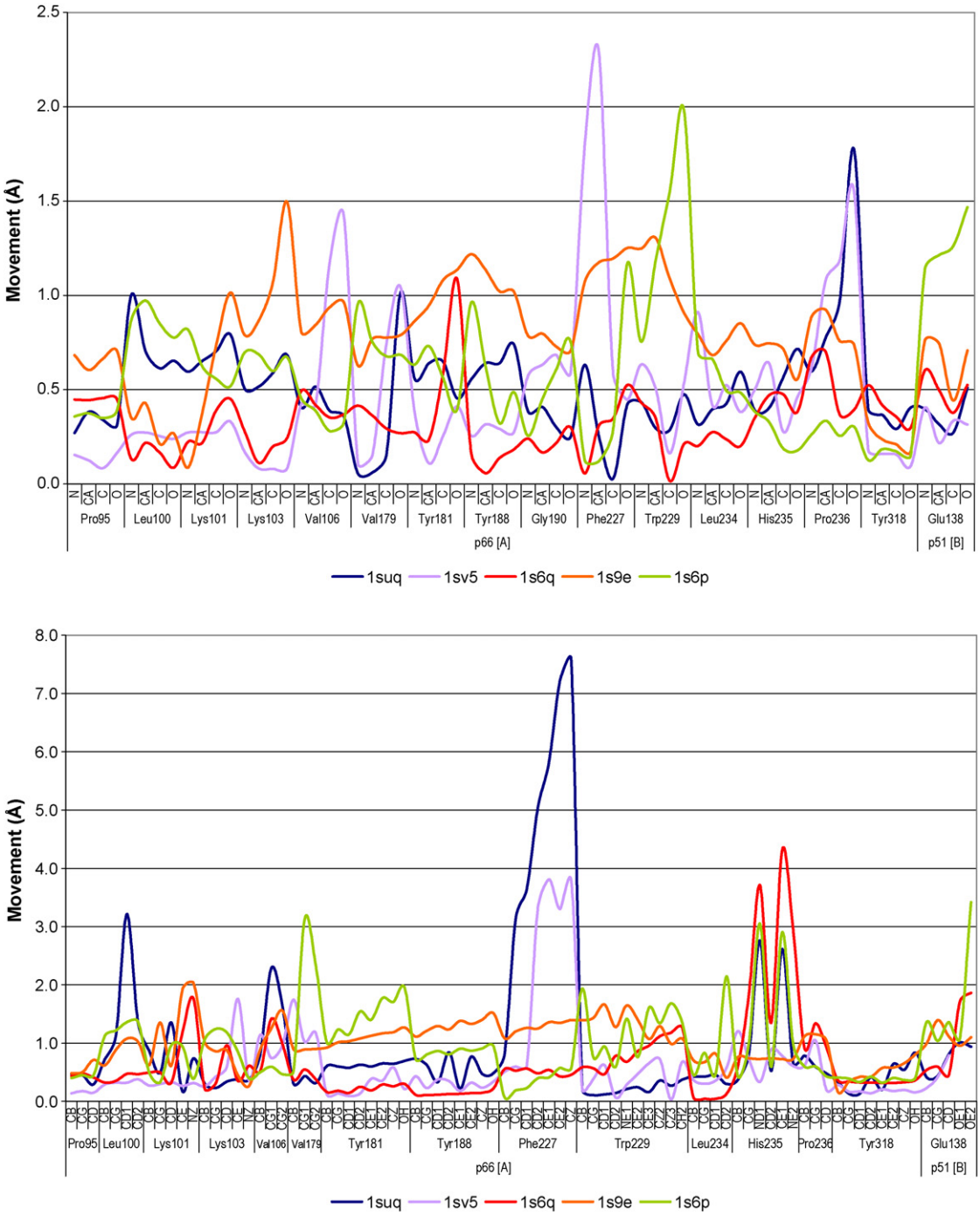


Fig. 5. Identification of the scalar change in the position of individual backbone atoms (top graph) and side chain atoms (bottom graph) of the NNIBP upon relaxation of the crystal structures.

dynamics simulations. The conformation of each ligand is excellent, but expected considering the rigidity of these particular inhibitors. The position of each ligand is close (average atom displacement for **4** (1.25 Å) and **5** (1.68 Å)), with a downward twist towards Tyr181 and Tyr188 observed for both inhibitors, with Tyr318 accommodating these changes with a sympathetic move downwards. There is some reorientation of the aliphatic sidechains as has been previously observed in the relaxation studies.

3.8. Ligand exchange

3.8.1. Prediction of ligand conformation in RT mutants

Crystal structures are available of nevirapine and efavirenz bound with RT with numerous single-point mutations in the binding pocket. A clinically important mutation occurs at Lys103Asn, in which nevirapine exhibits five-fold less binding affinity when compared to native RT, and has been crystallized (1fkp) [15]. To simulate the induced fit of a ligand in a mutated

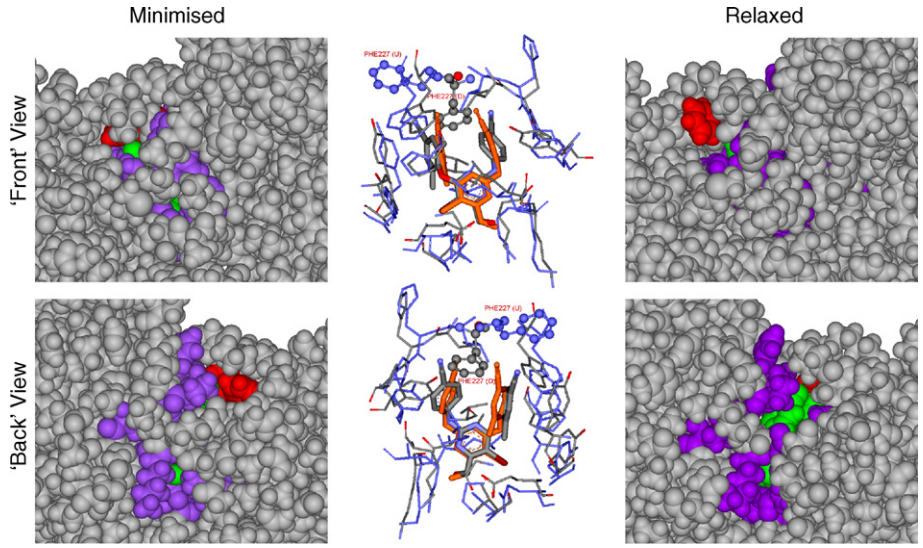


Fig. 6. The 'front' and 'back' views of RT (1suq) showing the different orientations of Phe227 (red space filling). The NNIBP (purple space filling), the ligand (green space filling) and the rest of the protein (grey space filling) are also shown. The superimposed atoms of the binding pocket (coloured blue-*relaxed*, atom type-*minimized*, Phe227 in ball and stick) and ligand (coloured orange-*relaxed*, atom type-*minimized*) are shown in the centre with hydrogens omitted for clarity.

protein, nevirapine **4** was sourced from 1vrt in its native binding mode and was superimposed into the crystal structure of 1sv5, the Lys103Asn mutant structure that originally contained the bound inhibitor **3**. The methodology (Scheme 1, path B) was applied, and the *relaxed* structure was then compared to the *minimized* structure of 1fkp (nevirapine bound to Lys103Asn RT mutant) (Fig. 8) by superimposition. The RMSD of superimposition was 1.42 Å, with an average atom displacement of the ligand of 1.98 Å. A comparable experiment was also performed with efavirenz **5**, sourcing the ligand from 1fk9 (native) and comparing the result to 1fko (Lys103Asn mutant) with similar results observed.

Unlike the pool of crystal structures used for the relaxation study, the mutant crystal structures 1fko and 1fkp had a different orientation of the Trp229 sidechain compared to their respective ligand bound native structures and the source of RT for the exchange (1sv5). This difference explains the relatively poor fit in these simulations. As the Trp229 indole sidechain

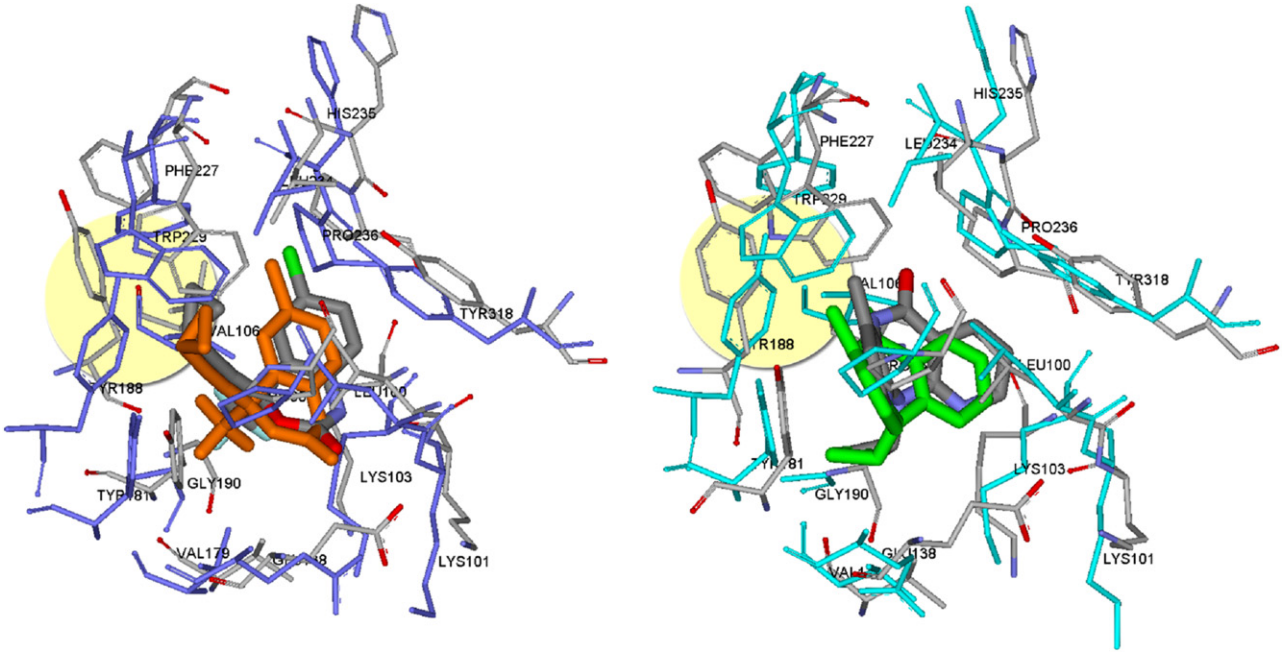


Fig. 7. Superimpositions from the simulation of the exchange of ligands to recreate their native binding modes. The minimized structure is indicated by element colour in both images. Left: the result from the ligand exchange of efavirenz **4** (1fk9) is shown in purple (NNIBP) and orange (ligand) sticks. Right: the result from the ligand exchange of nevirapine **5** (1vrt) is shown in aqua (NNIBP) and green (ligand) sticks. The slight distortions of Tyr188 and Trp229 are highlighted by the yellow circles in both images. Hydrogen atoms are omitted for clarity.

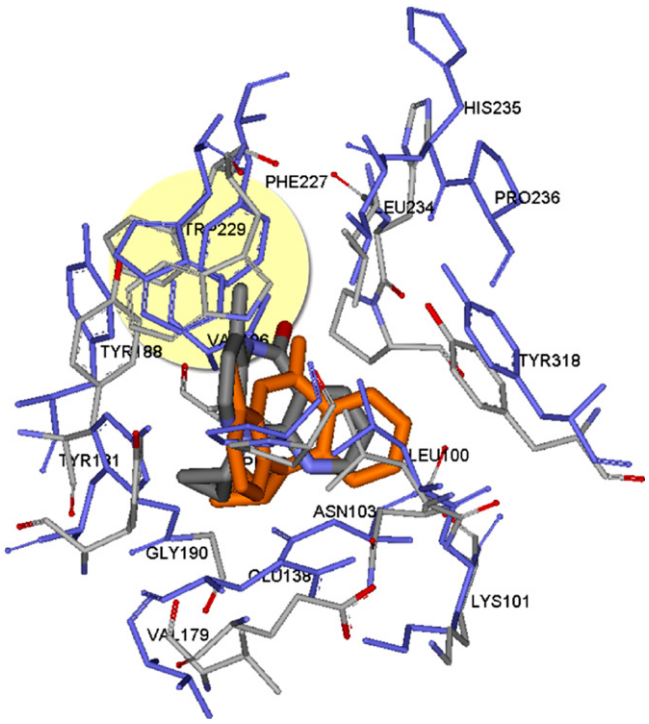


Fig. 8. Superimposition experiment attempting to recreate the induced fit of a ligand into RT with the Lys103Asn point mutation. The minimized structure of 1fkp is shown by element colour (RT NNIBP and ligand). The relaxed structure of the ligand from 1vrt superimposed into 1sv5 is shown in purple (RT NNIBP) and orange (nevirapine) sticks. The different orientations of Trp229 are highlighted by the yellow circle. Hydrogen atoms are omitted for clarity.

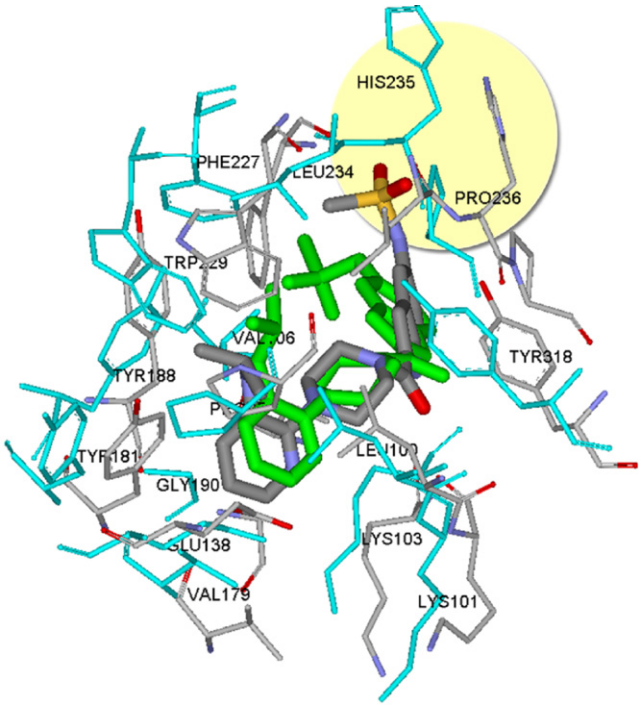


Fig. 9. Superimposition of the result of the ligand exchange of **6** into 1s6p onto the minimized structure of 1klm. The minimized structure is shown in sticks by element colour. The result from the ligand exchange is shown in aqua (NNIBP) and green (ligand) sticks. The large distortion of His234 and Pro235 is highlighted by the yellow circle. Hydrogen atoms are omitted for clarity.

still maintains one of the important π-stacking orientations with the ligand it is unlikely to 'flip' in the timescale of our simulations. The original source [16] does not propose why these differences have occurred or if they are artefacts of the data collection, or how important this is for the function of the mutated RT.

3.9. Ligand exchange

3.9.1. Induced fit of very large ligand

The limits of the present methodology were determined by attempting to simulate the interaction of a known ligand which would have a poor initial fit. The ligand delavirdine **6** (1klm) is a large ligand (Fig. 1) with a different binding mode to those investigated so far (Fig. 9). The ligand is partially exposed to the solvent from the back of the NNIBP between Val106 and His236. Initial superimposition of the ligand into the crystal structure of 1s6p exhibited significant bad contacts between the ligand and residues His235-Pro236. Application of the simulation protocol (Scheme 1, path B) produced a substantial reordering of the NNIBP with moderate agreement to the binding mode in 1klm. The superimposition of the *relaxed* structure with the *minimized* crystal structure had an RMSD of 1.91 Å. The major distortions of the binding pocket are localized at the position where the ligand showed bad contacts with the binding pocket, with a concomitant reorientation of Tyr181 and Tyr188 and rotation of Phe227. The rotation upwards of the dimethylanilino group of the ligand has resulted in a significant distortion of Trp229 (part of the primer grip). There was also some rearrangement of the sidechains due to a slight shift of the ligand out of the binding pocket. The movement of the inhibitor out of the pocket may have had a significant effect on the inhibitor's ultimate orientation in the pocket.

4. Experimental

4.1. Computational requirements

Molecular modelling was performed on an Intel Pentium 4 workstation (3.00 GHz processor and 2.00 GB of RAM) running under Windows XP (Service Pack 2). The proteins were manipulated with DS Modeling 1.1 (Accelrys®) or DS ViewerPro 5.0 (Accelrys®) [17]. Difference distance matrices were calculated with Proflex 2.0, a software package developed in-house (available upon request) [7]. Average run time for a molecular dynamics protocol was 48 h.

4.2. Crystal structure selection criteria

All available HIV-1 RT files were downloaded from the RCSB PDB [18]. Nine structures were excluded from this study as they had a crystal structure resolution >3.0 Å. Three crystal structures were excluded (3hvt, 1hni, 1hnv) as a significant number of residues had been modelled as alanine during refinement. Three structures were excluded (1klm, 1rev, 1rtj)

because they resulted from the soaking out of a ligand which was then replaced by another. Comparisons using Proflex 2.0 indicated that these structures were essentially identical to the crystal structure from which the ligand was soaked out.

Crystal structure 1dlo could not be excluded even though it contained residues that were treated as alanine during refinement. This was the highest resolution crystal structure that was available of the unliganded (native) enzyme. It was chosen over 1hmv or 1rtj due to its higher resolution and data completeness, and because of the known irregularities in the crystal collection method of 1rtj [19]. This resulted in a set of 67 structures suitable for analysis.

4.3. Immobile atom (superimposition) subset determination

The immobile atom (superimposition) subset was created using Proflex 2.0. Briefly, DDMs were calculated in an automated two-step process. Firstly, an $n \times n$ square distance matrix (DM), where n is the number of residues in the protein, is constructed by calculating all distances r_{ij} between the C_α atoms of the ith and jth amino acids.

Separate DMs are calculated for two protein conformations, A and the comparison protein B, containing the distances r_{ij}^{A} and r_{ij}^{B}, respectively. Subtraction of one DM from the other, by calculating the difference Δr_{ij} between elements r_{ij}^{A} and r_{ij}^{B} that have corresponding values of i and j, yields the DDM comparing the conformations of A and B. If the two conformations A and B are identical, even if not superimposed, the corresponding distances r_{ij}^{A} and r_{ij}^{B} will be identical, and so the difference Δr_{ij} is equal to zero. However, if a residue x is in a different position in one structure it will have different distances to every other residue j in the protein and the differences in the DDM will be nonzero for every Δr_{ij} value where $i = x$. Thus, regions of dissimilarity between the two structures will appear as clusters of nonzero difference values in the DDM, with the value of Δr_{ij} indicating the magnitude of conformational change. DDMs can also be calculated for all atoms for any part of the protein to be compared.

Each crystal structure was compared to 1dlo, the unliganded native structure of HIV-1 RT. The 66 crystal structures were aligned to 1dlo using *align homologous sequence*, comparing C_α atoms only. The subset of atoms that differed only by Δr_{ij} 0–2 Å was calculated using the default threshold settings (determined on the total number of points within the DDM that are within the subset range to be analysed, as a percentage of the total number of points within the matrix). The 66 subsets were then merged with Proflex 2.0 to find the atoms common to all subsets producing a superimposition subset of 318 C_α atoms. Following comparison of this subset with the collection method data of 1dlo, an additional 6 C_α atoms were excluded as they were treated as alanine in the original 1dlo refinement: these residues were p66 Subunit; Glu194, Glu203, Gln407, Trp414, p51 subunit; Glu36, and Gln407. Additionally, the 10 C_α atoms of the NNIBP that appeared in the superimposition subset were also removed. The NNIBP was defined as Chain A [p66]: Pro95, Leu100, Lys101, Lys103, Val106, Val179, Tyr181, Tyr188, Gly190, Phe227, Trp229, Leu234, His235, Pro236, Tyr318 and Chain B [p51]: Glu138 [14]. While Proflex 2.0 calculations were performed on all C_α atoms, the dynamics protocol employed (*vide infra*) places constraints on all the atoms of a residue, but no constraints were placed on any atoms inside the binding pocket. Therefore, the superimposition subset, i.e. C_α atoms that only moved between 0 and 2 Å on the binding of any substrate (defined in the experimental as the 0–2_Subset), was 302 C_α atoms (comprising 91 from the p66 subunit (Chain A), and 211 from the p51 subunit (Chain B)), and is listed below.

- *p66 subunit (chain A)*. 96–99, 104–105, 107, 162–163, 166–167, 180, 189, 191–193, 196, 198–200, 202, 319–321, 323–329, 339–345, 349–351, 366, 368, 370–395, 397–402, 405–406, 408–411, 415–416, 428, 431–435, 493 and 530.
- *p51 subunit (chain B)*. 7, 11, 19, 21, 23–35, 37–50, 57–63, 71–81, 84, 97, 99–100, 105–120, 122, 124–137, 139–153, 155–158, 161–162, 164–165, 167–170, 178–189, 234–235, 247, 254–255, 263, 266–267, 290, 293, 303, 323–334, 336–344, 347–354, 363–393, 398–406 and 408–415.

4.4. Molecular dynamics simulation procedure

4.4.1. Structure preparation

Each structure was imported into DS Modeling 1.1 using the Clean option (Fix incomplete or incorrect residues, add termini (pH 7), and add missing hydrogens). Waters of crystallization were deleted (1klm, 1s6p, 1vrt). The bond order of the ligand was corrected (introducing aromaticity, double bonds, etc.) and then hydrogens were added to the ligand. Each structure was atom typed using the Accelrys® CHARMm forcefield [20].

4.4.2. Protocol—crystal structure relaxation

Each crystal structure was subjected to a stepwise minimization and relaxation using the CHARMm (Version 29b1) [21] molecular dynamics module as implemented in DS Modeling 1.1. The system was an *in vacuo* simulation using default non-bonding parameters (cutoff: 13.50; cutoff-on: 8.00; cutoff-off: 12.00). The structure was minimized with Steepest Descent for 3000 steps (Grad. Tol. 0.1) with all non-hydrogens explicitly fixed (i.e. minimization of all hydrogens). The structure was minimized with Steepest Descent for 3000 steps (Grad. Tol. 0.1) with no constraints. The structure was minimized with Adopted Basis-set Newton–Raphson (ABNR) for 30,000 steps (Grad. Tol. 0.01) with no constraints. The output structure was designated the *minimized* structure. The gradient tolerance was required to converge during this step. The system was heated to 300 K over 2,000 steps (all atoms of 0–2_Subset explicitly fixed), then equilibrated for 10,000 steps (0–2_Subset explicitly fixed). Molecular dynamics (NVT) (i.e. production) was run at 300 K, for 175,000 steps (0.175 ns) with data saved every 1000 frames (0–2_Subset explicitly fixed).

Early in the method development the system was minimized at this point using steepest descent (Grad. Tol. 0.1) for 2000 steps, and then ABNR for 30,000 steps (Grad. Tol. 0.01) with no constraints. Convergence was reached during this step and the resulting structure was designated the *relaxed* structure.

Later methodology development required the frame with the minimum energy from the production step to be identified with the Analysis module. The same two-step minimization was applied on this frame, as above, until convergence was reached to achieve the *relaxed* structure.

4.5. Ligand exchange simulations

4.5.1. Protocol—exchange with similar ligands

The ligand exchange simulations were performed using *minimized* crystal structures (the output structure from the first three steps of the relaxation protocol). The crystal structure from which the ligand was sourced was superimposed on the second crystal structure using the superimposition by alignment of defined residues as implemented in DS Modeling 1.1, using the superimposition subset determined with Proflex 2.0. The ligand was copied from the transformed crystal structure and pasted into the model system and the original ligand was deleted. The system was then retyped. Due to a reported bug in DS Modeling 1.1, structures are required to be retyped when the file has been modified, even if they have been previously typed. Before this is performed it is critical that from the *forcefield typing options* dialogue box the *fix peptide and crystallographic water* option is deselected, otherwise, when the structure is retyped, the aromatic NH protons of all the histidine and tryptophan residues will be deleted.

The dynamics simulation was a seven-part stepwise protocol. The system was minimized by steepest descent for 3000 steps (Grad. Tol 0.1) with all the atoms of RT explicitly fixed so the ligand could adjust for bad contacts arising from the superimposition. The system was then minimized by ABNR for 30,000 steps (Grad. Tol. 0.01) with no constraints. The gradient tolerance was required to converge during this step. The system was heated to 300 K over 2000 steps (0–2_Subset explicitly fixed), then equilibrated for 10,000 steps (0–2_Subset explicitly fixed). Molecular dynamics (NVT) was run at 300 K, for 175,000 steps (0.175 ns) with data saved every 1000 frames (0–2_Subset explicitly fixed). The frame with the minimum energy was then minimized using steepest descent (Grad. Tol. 0.1) for 2000 steps, and then ABNR for 30,000 steps (Grad. Tol. 0.01) with no constraints.

4.5.2. Protocol—exchange with dissimilar ligands

The crystal structure from which the dissimilar ligand was sourced was *minimized* as described previously. Note that while these structures contained incomplete domains, no attempt was made to reconstruct missing residues. The intention was only to minimize the structure with hydrogens added in the new forcefield, the previous results indicated that there would likely be no significant deviations from the crystal structure. The ligand was then superimposed onto *minimized* 1s6p as described previously and the seven-part stepwise protocol applied as described for the exchange with similar ligands.

4.5.3. Protocol—ligand exchange into RT mutant

The crystal structures from which nevirapine was sourced (1vrt, native) and compared (1fkp, Lys103Asn mutant) were both *minimized* as described previously. The ligand was then superimposed into *minimized* 1sv5 as described previously and the seven-part stepwise protocol applied as described for the exchange with similar ligands. The *relaxed* structure was compared by superimposition to *minimized* 1fkp.

5. Discussion

Our newly developed method for the simulation of induced fit between a ligand and its target protein is based upon a desire to account for complete protein and ligand flexibility, while at the same time restricting computational effort. We achieve this balance by taking into account the wealth of information available from crystal structures of the protein under study in different conformations.

Using the example of the HIV-1 reverse transcriptase (RT) enzyme, we have analysed 67 crystal structures in different conformations with our difference distance matrix (DDM) technique. We confirmed our previous finding [7] with a smaller set of crystal structures, that approximately one third of this large protein (nearly 1000 amino acids) remains immobilized, regardless of whether substrate analogues, or competitive or allosteric inhibitors, or none of these are complexed to the protein.

It therefore seemed reasonable to fix those immobile residues during our molecular dynamics (MD) simulations so as to avoid computational effort being expended on those parts of the protein which are highly unlikely to move. For the same reason of minimizing computational effort, simulations have been performed *in vacuo*. This was considered reasonable for the very hydrophobic pocket of RT, where water molecules are not always crystallised, and the ones that are present do not cluster in defined regions in space. If dealing with a less hydrophobic pocket, consideration should be given to the inclusion of explicit solvation of the binding pocket or the use of an implicit solvent model.

To validate our method, we needed firstly to subject a number of crystal structures with complexed non-nucleoside inhibitors to our constrained simulation protocol to ensure they would not be distorted.

Results showed that average atom movements were small (Table 1) and fast convergence of energy and changes in geometry (as measured by RMSD) were observed. Even though the atoms of the non nucleoside inhibitor binding pocket (NNIBP) were excluded from the set of constrained atoms, we found that they moved little during the MD simulations. This highlights the stabilization of the binding site by the ligand, and also shows that fixing one third of the protein does not result in distortions during MD simulations.

Interestingly, the crystal structure with the ligand of the lowest affinity of the five structures tested showed the largest deviation from the original crystal structure after the simulations.

The next step in the development of the new constrained MD method resulted in successful ligand exchanges between crystal structures. The fit of structurally dissimilar inhibitors, such as nevirapine and efavirenz (Fig. 1), could be simulated

satisfactorily. In particular it was encouraging to be able to observe the same interactions between the ligands and the RT after the simulations as those in the original crystal structures.

In accordance with the findings for crystal structure relaxation, it proved more difficult to simulate the fit of nevirapine into a mutant enzyme, to which nevirapine binds with reduced affinity.

Finally, the simulation of the fit of a large ligand with a different binding mode was attempted and proved moderately successful.

6. Conclusion

Our new methodology of determining suitable constraints and then applying these in MD simulations has been shown to be robust in not introducing distortions. We have been able to simulate the induced fits of ligands into preformed binding pockets. Short simulation times are sufficient to achieve good fits. This makes this method ideally suited to the refinement of results from automated docking procedures, with only a small computational effort. We are further developing our constrained MD methodology for simulating induced fit to overcome its limitations. We are currently restricted to modelling ''well-fitting'' ligands, that is, those that show high affinity and are not too large to fit into the preformed pocket. In order to model successively larger changes in protein conformation, we propose to introduce a more complex environment of constraints of varying types and strengths around the binding pocket, which will allow more flexibility, but will necessitate longer simulation times.

Acknowledgements

This work is supported by the Australian Research Council (LP0348732) and Avexa Ltd.

References

[1] M.B.B. Samir Elamrani, G.N. Phillips Jr., J.A. McCammon, Study of global motions in proteins by weighted masses molecular dynamics: adenylate kinase as a test case, Proteins: Struct., Funct., Genet. 25 (1996) 79–88.

[2] H.A. Carlson, Protein flexibility and drug design: how to hit a moving target, Curr. Opin. Chem. Biol. 6 (2002) 447–452.

[3] H.A. Carlson, J.A. McCammon, Accommodating protein flexibility in computational drug design, Mol. Pharmacol. 57 (2000) 213–218.

[4] S.J. Teague, Implications of protein flexibility for drug discovery, Nat. Rev. Drug. Discov. 2 (2003) 527–541.

[5] K.L. Meagher, H.A. Carlson, Incorporating protein flexibility in structure-based drug discovery: using HIV-1 protease as a test case, J. Am. Chem. Soc. 126 (2004) 13276–13281.

[6] I.L. Alberts, N.P. Todorov, P.M. Dean, Receptor flexibility in de novo ligand design and docking, J. Med. Chem. 48 (2005) 6585–6596.

[7] P.A. Keller, S.P. Leach, T.T.T. Luu, S.J. Titmuss, R. Griffith, Development of computational and graphical tools for analysis of movement and flexibility in large molecules, J. Mol. Graph. Model. 18 (2000) 235–241.

[8] K. Nishikawa, T. Ooi, Y. Isoga, N. Saito, Tertiary structure of proteins I: representations and computation of the conformation, J. Phys. Soc. Jpn. 32 (1992) 1331–1337.

[9] For examples of other defined superimposition subsets see A.C.W. May, Pairwise iterative superimpositions of distantly related protein and assessment of the significance of 3-D structural similarity, Protein Eng. 9 (1996) 1093–1101.

[10] M. Akke, S. Forsen, W.J. Chazin, Solution structure of $(Cd^{2+})_1$-calbindin D_{9k} reveals details of the stepwise structural changes along the Apo $\rightarrow (Ca^{2+})_1^{II} \rightarrow (Ca^{2+})_2^{I,II}$ binding pathway, J. Mol. Biol. 252 (1995) 102–121.

[11] R.M. Esnouf, J. Ren, C. Ross, Y. Jones, D. Stammers, D. Stuart, Mechanism of inhibition of HIV-1 reverse transcriptase by non-nucleoside inhibitors, Nat. Struct. Biol. 3 (1995) 303–308.

[12] J. Ding, K. Das, Y. Hsiou, S.G. Sarafianos, A.D.J. Clark, A. Jacobo-Molina, C. Tantillo, S.H. Hughes, E. Arnold, Structure and functional implications of the polymerase active site region in a complex of HIV-1 RT with a double-stranded DNA template-primer and an antibody Fab fragment at 2.8 Å resolution, J. Mol. Biol. 284 (1998) 1095–1111.

[13] K. Das, A.D. Clark, P.J. Lewi, J. Heeres, M.R. deJonge, L.M.H. Koymans, H.M. Vinkers, F. Daeyaert, D.W. Ludovici, M.J. Kukla, B. DeCorte, R.W. Kavash, C.Y. Ho, H. Ye, M.A. Lichtenstein, K. Andries, R. Pauwels, M.-P. deBethune, P.L. Boyer, P. Clark, S.H. Hughes, P.A.J. Janssen, E. Arnold, Roles of conformational and positional adaptability in structure-based design of TMC125-R165335 (Etravirine) and related non-nucleoside reverse transcriptase inhibitors that are highly potent and effective against wild-type and drug-resistant HIV-1 variants, J. Med. Chem. 47 (2004) 2550–2560.

[14] J.S. Ren, R.M. Esnouf, E. Garmen, D. Somers, C. Ross, I. Kirby, J. Keeling, G. Darby, Y. Jones, D. Stuart, D. Stammers, High resolution structures of HIV-1 RT from four inhibitor complexes, Nat. Struct. Biol. 2 (1995) 293–302.

[15] G. Maga, M. Amacker, N. Ruel, U. Hubscher, S. Spadari, Resistance to nevirapine of HIV-1 reverse transcriptase mutants: loss of stabilizing interactions and thermodynamic or steric barriers are induced by different single amino acid substitutions, J. Mol. Biol. 274 (1997) 738–747.

[16] J. Ren, J. Milton, K.L. Weaver, S.A. Short, D.I. Stuart, D.K. Stammers, Structural basis for the resilience of Efavirenz (DMP-266) to drug resistance mutations in HIV-1 reverse transcriptase, Structure 8 (2000) 1089–1094.

[17] DS Modeling 1.1, Accelrys, San Diego, CA, USA, 2002.

[18] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, et al., The protein data bank, Nucl. Acids Res. 28 (2000) 235–242. See: http://www.rcsb.org/pdb/index.html.

[19] S.G. Sarafianos, K. Das, J. Ding, Y. Hsiou, S.H. Hughes, E. Arnold, Current perspectives on mechanisms of HIV-1 reverse transcriptase inhibition by nonnucleoside inhibitors, Spec. Publ.: R. Soc. Chem. 198 (1997) 328–334.

[20] F.A. Momany, R. Rone, Validation of the general purpose QUANTA 3.2/CHARMm force field, J. Comput. Chem. 13 (1992) 888–900.

[21] B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, M. Karplus, CHARMM: a program for macromolecular energy, minimization, and dynamics calculations, J. Comput. Chem. 4 (1983) 187–217.