ORIGINAL PAPER

Comparative and pharmacophore model for deacetylase SIRT1

Tero Huhtiniemi · Carsten Wittekindt ·
Tuomo Laitinen · Jukka Leppänen ·
Antero Salminen · Antti Poso ·
Maija Lahtela-Kakkonen

Abstract Sirtuins are NAD-dependent histone deacetylases, which cleave the acetyl-group from acetylated proteins, such as histones but also the acetyl groups from several transcription factors, and in this way can change their activities. Of all seven mammalian SirTs, the human sirtuin SirT1 has been the most extensively studied. However, there is no crystal structure or comparative model reported for SirT1. We have therefore built up a three-dimensional comparison model of the SirT1 protein catalytic core (domain area from residues 244 to 498 of the full length SirT1) in order to assist in the investigation of active site–ligand interactions and in the design of novel SirT1 inhibitors. In this study we also propose the binding-mode of recently reported set of indole-based inhibitors in SirT1. The site of interaction and the ligand conformation were predicted by the use of molecular docking techniques. To distinguish between active and inactive compounds, a post-docking filter based on H-bond network was constructed. Docking results were used to investigate the pharmacophore and to identify a filter for database mining.

Tero Huhtiniemi and Carsten Wittekindt contributed equally to this work.

T. Huhtiniemi (✉) · C. Wittekindt · T. Laitinen ·
J. Leppänen · A. Poso · M. Lahtela-Kakkonen
Department of Pharmaceutical Chemistry, University of Kuopio, P.O. Box 1627, Kuopio FIN-70211, Finland
e-mail: Tero.Huhtiniemi@uku.fi

A. Salminen
Department of Neuroscience and Neurology, University of Kuopio, P.O. Box 1627, Kuopio FIN-70211, Finland

Introduction

The human histone deacetylases (HDACs) can be separated into three classes: class I contains HDACs 1–3 and 8, class II contains HDACs 4–7 and 9–10, and class III contains seven evolutionarily conserved Sir2 family members or Sirtuins SirT1–7 [1]. Histone deacetylases are enzymes which cleave the acetyl-group from acetylated proteins, such as histones but also the acetyl groups from several transcription factors, and in this way can change their activities. Several recent reports have indicated that Sirtuins are NAD-dependent histone deacetylases [2–4]. Sirtuins couple protein substrate deacetylation and NAD^+ cleavage to produce 2′- and 3′-O-acetyl-ADP-ribose [3], a novel product whose biological function is still unclarified. The NAD^+-dependency of Sirtuin type of HDACs provides an interesting link between energy metabolism and gene regulation. Leonard Guarente [2] has proposed that Sirtuins (SirTs) might be the link between metabolic activity, genomic stability and the aging process.

Of all seven mammalian SirTs, SirT1 has been the most extensively studied [5]. It is highly expressed in several adult tissues, such as brain, heart and skeletal muscle. SirT1 binds several important transcription

factors, such as p53, p65 (NF-kB), MyoD, and regulates their activity [6]. For instance, the deacetylation of p53 prevents the cells from undergoing apoptosis [7]. SirT1 also increases the ability of FoxO3 to induce cell cycle arrest and increases cellular resistance to oxidative stress [8]. Furthermore it interacts with PGC-1α, which is a transcriptional co-activator and a major metabolic regulator [9] and that way SirT1 can regulate cellular metabolism, for instance, mitochondrial biogenesis and glucose homeostasis [9, 10].

The biological functions of SirT1 have triggered interest in the development of SirT1 activators and inhibitors. Some activators have been reported, e.g. resveratrol, fisetin and butein, which are plant polyfenols [11–13]. Inhibition of SirT1 has been observed

after treatment at the low μM level, with nicotinamide, sirtinol or splitomicin (Fig. 1) [14–18]. Recently, a more potent inhibitor was described EX527 [19, 20], which is approximately 1000-fold more potent than nicotinamide.

There is structural conservation among the protein family of sirtuins. In all organisms ranging from bacteria to humans, the sirtuin protein family contains a core domain that consists of a series of conserved sequence motifs [21]. The catalytic core (Figs. 2, 3) consists of an NAD^+-binding domain in a variation of the Rossmann fold and of a smaller sub-domain composed of a helical module and a zinc-binding module. A large groove at the interface between these domains creates the active site where NAD^+ and the acetylated substrate bind.

The acetyl lysine substrate has been proposed to bind in a narrow channel that terminates near to the nicotinamide ribose of NAD^+. Binding of acetyl-substrate is believed to mediate bending of the nicotinamide ring of NAD^+ in a strained conformation also referred to productive NAD^+ binding, promoting the cleavage of the ribosyl–nicotinamide bond. Subsequently there is an attack on the acetyl group facing ribose to yield 2′-OAADPr and the deacetylated substrate-lysine [23–25].

The aim of this present study was to build up a comparison model of the human sirtuin SirT1 protein catalytic core (domain area from residues 244 to 498 of the full length SirT1) in order to assist in the investigation of active site–ligand interactions and in the design of novel SirT1 inhibitors. Molecular dynamic simulations and ligand docking were used to refine and
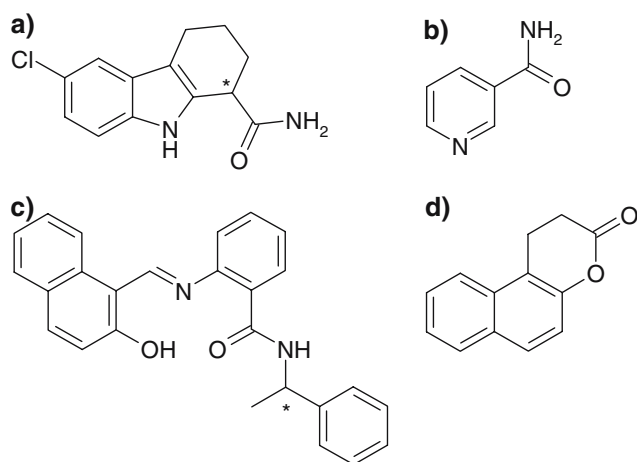
Fig. 1 SirT1 inhibitors: (a) EX527 (b) Nicotinamide (c) Sirtinol (d) Splitomicin. (The asterisks denote chiral centres)

Fig. 2 SirT1 model with the inhibitor EX527 (compound 1(S)) docked at the interface of a Rossmann fold-like large sub-domain (on the right) and the small sub-domain (on the left). The surface of the protein is coloured brown for lipophilicity and blue for the hydrophilic areas. Atoms of the inhibitor highlighted in blue have hydrogen bond donors binding to the polar surface. The aromatic and aliphatic parts settle on the lipophilic area. The "Flexible loop" above the ligand in the picture restricts the size of the amide end binding cavity
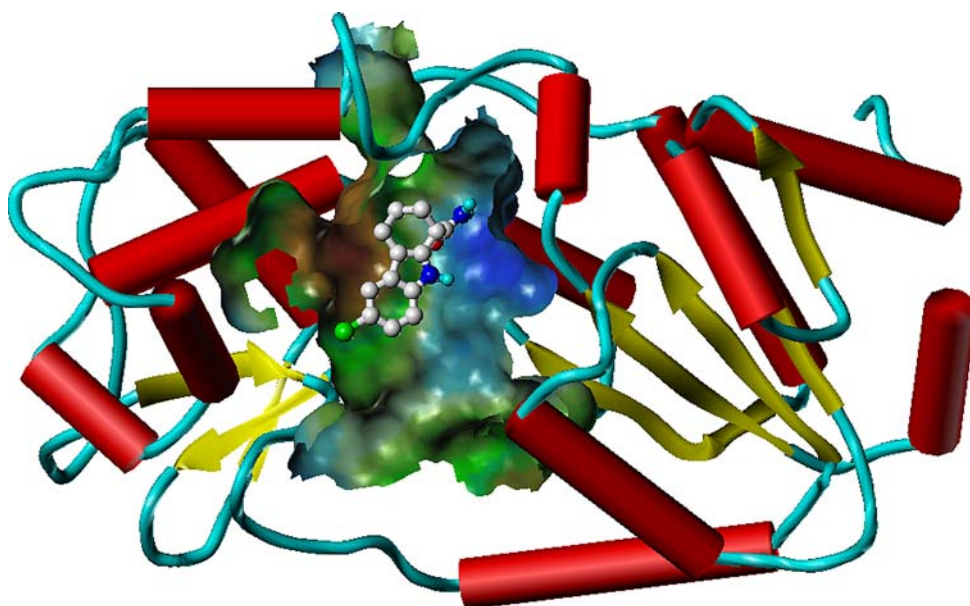
Fig. 3 Secondary structure of modeled SirT1 and reported [22] SirT2 structure. The SirT1 coordinates for all β-sheets coloured in yellow are assigned directly from the SirT2 structure. Helices are coloured in magenta and loops in cyan. Dash line represents the interface between two sub-domains
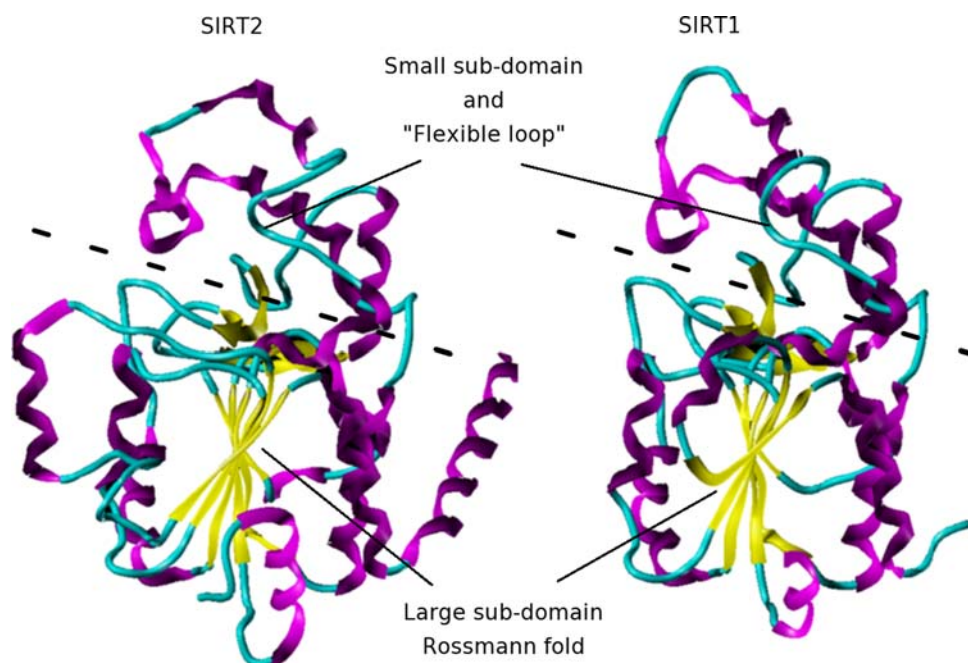
validate the model. In this study we also report the proposed binding-mode of EX527 in SirT1. An extensive H-bond network between EX527 and SirT1 in the binding site of ribose-nicotinamide complex of NAD$^+$ is proposed as a pharmacophore, preventing the NAD$^+$-dependent deacetylation mechanism in SirT1.

Experimental section

Methods

SirT1 domain sequence (Q96EB6 amino acids 244–498) was retrieved from the SWISS-PROT protein sequence database. Template sequences and structures were retrieved from the RCSB Protein Databank [26]. Blast [27] and SAS [28] scoring were used for selecting those template sequences producing significant alignments with the SirT1 domain. The multiple sequence alignment, comparative modelling and energy minimization were carried out using the Insight II [29] molecular modelling package. For comparative modelling, the HOMOLOGY module implemented in Insight II was used.

The protein structure was energy minimized using cvff forcefield as implemented in Insight II. Molecular dynamics (MD) simulations were performed using the GROMACS package [30, 31]. The stereochemical quality of the resulting protein structure was tested with the PROCHECK [32] computer program.

Molecular interaction field (MIF) calculation was performed using GRID version 1.2.2 [33]. The docking of the SirT1 inhibitor set was performed with the genetic algorithm based program GOLD [34]. Analysis of the interaction between ligand and protein were undertaken with the SILVER tool of the GOLD software package. The molecular structures of the ligands were generated using the Sybyl 7.1 software package [35] and geometries were energy minimized using Tripos force field without charges.

Comparative modelling

The crystal structures of the sirtuin deacetylase family of enzymes were used as templates in this study: Human SirT2, (1J8F.pdb) [22], *Saccharomyces cerevisiae* Hst2, (1Q1A.pdb) [36], *Thermotoga maritima* Sir2Tm, (1YC5.pdb) [37], *Archaeoglobus fulgidus* Sir2-Af2, (1S7G.pdb) [25], *Archaeoglobus fulgidus* Sir2-Af1, (1M2G.pdb) [38], *Eschericia coli* cobB, (1S5P.pdb) [39]. The resolution of the crystal structure was one criterion for the selection. Also the templates with mutated sites were discarded. Resolutions of the selected crystals were below 2.0 Å except for 1S7G resolution which was 2.30 Å. Template structures were selected for alignment due their BLAST and SAS scoring (supporting information). All six template sequences were aligned with SirT1 using blosum scoring matrix with gap penalties—gap_open 5, gap extension 0.05. (Align123 in HOMOLOGY). Alignment was

corrected and verified manually by superimposing structurally conserved regions and comparing their trace models (Fig. 4). Multiple alignment of 1J8F, 1Q1A, 1YCA, 1S7G, 1M2G and 1S5P sequences is shown in Fig. 5. The overall sequence identity of SirT1 is 44% with (1J8F), 38% with (1Q1A), 33% with (1YCA) while (1M2G) has 30% and (1S7G) has 29% of identical residues. Conservation among sequences highlighted in black and grey in the alignment is achieved with the GENEDOC [40] computer program.

SIRT2	HST2	SIR2Tm	SIR2AF2	SIR2AF1	COBB
1j8F	1Q1A	1YC5	1S7G	1M2G	1S5P
RMS	2.537378	2.522447	2.270012	2.221504	2.472396

Fig. 4 Conserved regions are superimposed well. RMS deviations of conserved regions are compared with SirT2 structure. Colour codes: black = 1J8F, yellow = 1Q1A, pink = 1YC5, blue = 1S7G, green = 1M2G, orange = 1S5P

In regions where the SirT1 sequence alignment with SirT2 was good, the backbone conformation was based on that of the SirT2 template structure. Where there were gaps in the alignment, other templates were used to generate possible backbone conformations due to their homology or RMS fitting. For loop areas where no conservation among sequences was found, the loop search method as implemented in Insight II utilizing the Protein Databank was used. Conformations for side-chains were evaluated by checking for improper steric overlaps between residues. This was done by rotating the residues to lowest non-bonding energy conformation starting from least conserved residues (non-coloured in the alignment).

The structure was minimized gradually (hydrogens → side chains → splice points → all) using constraints in heavy or backbone atoms to alleviate any remaining bad steric contacts. The final model was checked using the PROCHECK computer program, to search for deviations from normal protein conformational parameters.

Molecular dynamics

To refine side chain orientation, molecular dynamics (MD) simulations were performed on the structure. Therefore, the modelled structure was placed in a cubic water box at 300 K for 400 ps (after a 100 ps equilibration of the system with all the heavy atoms of the protein constrained). Particle mesh Ewald (PME) [41,

Fig. 5 Alignment of structures used for SirT1 comparative modelling. Conservation is highlighted using Genedock [40] multiple sequence alignment editor. Numbering under the alignment shows the sequences used for assigning coordinates for the SirT1 model. Residues marked with | are in the vicinity of the docked inhibitor EX527 (compound 1(S)). (Numbers and asterisks above the alignment denote the residue numbers)


```
                         |  |  |||||          |       | ||            |  |   |
                         *        20        *          40         *          60        *
1SirT1 : EDAVKLLQE--CKKIIVLTGAGVSVSCGIPDFRSR-DGIYARLAVDFPDLPDPQAMFDIEYFRKDPRPFFKFAKEI : 316
2SirT2 : EGVARYMQSERCRRVICLVGAGISTSAGIPDFRSPSTGLYDNLEKYH--LPYPEAIFEISYFKKHPEPFFALAKEL : 138
3Hst2  : RKIAAHMKSNPNAKVIFMVGAGISTSCGIPDFRSPGTGLYHNLARLK--LPYPEAVFDVDFFQSDPLPFYTLAKEL : 86
4Sir2Tm : KEFLDLLNE--SRLTVTLTGAGISTPSCGIPDF/--------------QNVFDIDFFYSHPEEFYRFAKEG : 67
5Sir2Af2 : RKAAEIIAA--SKHAVVFTGAGISAEGGIPDFRGE-DGLWRKYD--------PEEVASISGEKRNPRAFWEFSMEM : 70
6Sir2Af1 : EKLLKTIAE--SKYLVALTGAGVSAESGIPTFRGK-DGLWNRYR--------PEELANPQAFAKDPEKVWKWYAWR : 67
7CobB   : ----------KPRVLVLTGAGISAESGIRTFRAA-DGLWEEHR--------VEDVATPEGFDRDPELVQAFYNAR : 95
         222223555--5222222222222222222555555-5333222222222LLLLL2222222222222222222222222
```

```
           |  |||||      |          |
          80         *         100         *          120        *          140        *
1SirT1 : Y----PGQ--FQPSLCHKFIALSDKEGKL-LRNYTQNIDTLEQVAGIQ--RIIQCHGSFATASCLI--CKYKVDCE : 381
2SirT2 : Y----PGQ--FKPTICHYFIRLLKDKGLL-LRCYTQNIDTLERIAGLEQEDLVEAHGTFYTSHCVSASCRHEYPLS : 207
3Hst2  : Y----PGN--FRPSKFHYLIKLFQDKDVL-KRVYTQNIDTLERQAGVKDDLIIEAHGSFAHCHCIG--CGKVYPPQ : 153
4Sir2Tm : I--FPMLQ--AKPNLAHVLIAKLEEKGLI-EAVITQNIDRLHQRAGSK--KVIELHGNVEEYYCVR--CEKKYTVE : 134
5Sir2Af2 : KDK---LF--ABPNPAHYALAELERMGIV-KAVITQNIDMLHQRAGSR--RVLELHGSMDKLDCLD--CHETYDWS : 136
6Sir2Af1 : M--EKVFN--AQPNKAHQAFAELERLGVL-KCLITQNVDDLHERAGSR--NVIHLHCSLRVVRCTS--CNNSFEV- : 133
7CobB   : R--RQLQQPEIQPNAAHLAIAKLQDALGDRFLLVTQNIDNLHERAGNT--NVIHMHGELLKVRCSQ--SGQVLDW- : 164
         2----222--222222222222222222-2222222222444444--422222222222333--33222222
```

```
                                                 |
         160         *          180        *          200         *          220        *
1SirT1 : AVRGDIFN--QVVPRCPRCPADEPLAIMKPEIVFFGENLPEQFHRAMKYDKDE----------VDLLIVIGSSLK : 444
2SirT2 : WMKEKIFS---EVTPKCEDCQS-----LVKPDIVFFGESLPARFFSCMQSDFLK----------VDLLLVMGTSLQ : 265
3Hst2  : VYPKAIIEHPIKDFVKCDVCGEW-----LVKPAIVFFGENLPDSFSETWLNDSEWLREKITT/PQQP-VIVVGTSLA : 223
4Sir2Tm : DVIKKLES---SDVPLCDDCNS-----LIRPNIVFFGENLPDALREAIGLSSR--------ASLMIVLGSSLV : 192
5Sir2Af2 : EFVEDFNA---GEIPRCRACGSY----YVKPRVVLFGEPLPQRTLFEAIEEAKH----------CDAFMVVGSSLV : 195
6Sir2Af1 : --ESAPKI---PPLPKCDKCGS-----LLRPGVVWFGEPLPDVLDRAMREVER----------ADVIIVAGTSAV : 189
7CobB   : --TGDVTP----EDKC/---------PLRPHVVWFGEMPL--GMDEIYMALSM----------ADIFIALGTSGH : 211
         22222222---22222222222222222222222222222222222222222----------222222222222
```

```
            *           240         *         260          *         280          *         300
1SirT1 : VREVALIPSSIP-HEVPQILINREPLPHLHF---------------------DVELLGDCDVIINELCHRIGGEYA : 498
2SirT2 : VQPTASLISKAP-LSTPRLLINKEKAGQSDPFLGMIMGLGGGMDFDSKKAYRDVAWLGECDQGCLAIAELLGWKKE : 340
3Hst2  : VYPAASLPEEIP-RKVVKRVLCNLETVGDFKAN---------------KRPTDLIVHQYSDIFAEQLVEELLGWQED : 282
4Sir2Tm : VYPAAELPLITVRSGGKLVIVNLGETPFDDIA---------------TLKYNMDVVEFARRVMEEGGI--- : 245
5Sir2Af2 : VYPAAELPYIAKKAGAKMIIVNAEPTMADPIF---------------DVKIIGKAGEVLPKIVEEVKRLRS : 251
6Sir2Af1 : VQPAASLPLIVKQRGGAIIELINPDETPLTPIA---------------DYSLRGKAGEVLRHIVRHVRKALS : 245
7CobB   : VYPAAGFVHEAKLHGAHTVELINLEPSQ/--EF---------------AEKYYGPASCVVPEFVEKLLKGLK : 264
         22222222222-2222222222LLLLLLLL---------------------LL2222222222222222222
```

42] electrostatic and periodic boundary conditions were used as well as pressure and temperature coupling [43] to an external bath were used in the simulation. (See supporting information for detailed MD parameters). The backbone atoms were constrained with a force constant of 1000 kJ/(mol nm^2) so that only the side chains were able to move freely during the simulation. At the end of the run the side chains of the protein were minimized with the steepest descent method for 10,000 steps to remove bad contacts.

Ligand–receptor interactions

The model was prepared before docking by adding water molecules into a channel, which begins at the putative active site near the amino acids I347 and D348 and continues throughout the length of the entire enzyme [44]. The water was added in order to prevent the ligands docking into that channel. No direct interaction between the docked ligands and these water molecules was observed. The reported set of indole-based inhibitors [19] with IC$_{50}$ values ranging from the nanomolar to the micromolar levels, was docked at the binding site and the best docked configuration of each compound according to the GOLD fitness functions was taken for subsequent analysis. The analysis consisted of all possible H-bonds between ligand and protein as well as the occupancy of the space close to H363 by the ligands. The binding mode for the most potent compound was determined. To distinguish between active and inactive compounds, a post-docking filter based on H-bond network was constructed. Docking results were used to investigate the pharmacophore and to identify a filter for database mining.

The molecular interaction fields of the binding cavity for the modelled SirT1 were calculated using the GRID force field with all multi-atom-probes and all single-atom-probes, except metallic cations, 1 Å grid spacing and static protein treatment were applied. Hydrogens were added to the structure using the program GRIN. Those sites showing a higher interaction than water to the specific probe were evaluated as significant interaction sites.

Results and discussion

Comparative modelling

At first, the selected sequences of the protein family were aligned using the HOMOLOGY module tools of Insight II [29]. The alignment was corrected and verified manually by superimposing the structurally conserved regions and comparing their trace models with each other (Fig. 4). From this alignment (Fig. 5) it can be seen that the assignment of coordinates for most of the residues was quite straightforward, because the SirT2 structure could be used directly. Figure 6a illustrates the trace model of SirT1 superimposed on to the trace model of SirT2 before refining the structure. It can be seen clearly that in most areas the model has the same backbone structure as SirT2. Figure 3 depicts the secondary structures of both enzymes. All β-sheets and all helices sharing the same alignment have the backbone structure resembling that found in SirT2. For dissimilar areas in the alignment other homology structures were utilized for assigning coordinates or alternatively the structure was constructed by the loop
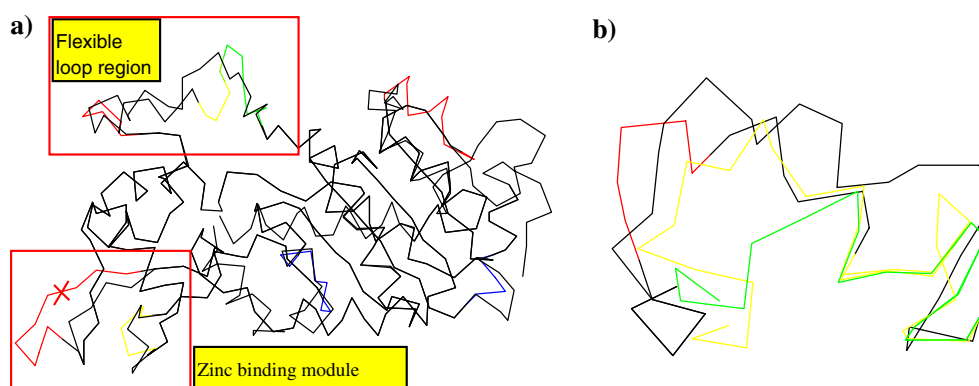
Fig. 6 (**a**) Trace model of SirT1 superimposed with SirT2. The black trace follows SirT2 structure. The coloured trace is according to the modelled SirT1 structure. The homology structures used are labelled with colour codes: black = 1J8F,

yellow = 1Q1A, blue = 1S7G, green = 1M2G, red = structure searched from RCSB Protein Databank. (**b**) Flexible loop region. Green and yellow traces of Sir2-Af1 and Hst2 reveal the closed orientation of the loop

search method implemented in Insight II utilizing the RCSB Protein Databank [26].

After minimising the structure, the model was subjected to MD simulation for 400 ps at 300 K keeping the backbone fixed. The MD simulation of the model was stable, as there was no drift detectable in the energy or temperature. The stereo-chemical quality of the model was checked by the average structure of the last 100 ps. The Ramachandran plot revealed only one residue in the disallowed region, seven in the generally allowed region and 17 in the additionally allowed regions. About 88% of all 190 residues were in the most favoured regions. Residue Leu-435 in the disallowed region was situated in a loop region and has therefore little significance. No residues in the disallowed region were situated in the active site.

Interface between large and small subdomain

The active site at the interface of Rossmann fold and small sub-domain can be seen in Figs. 2 and 3. The small sub-domain side of the interface consists of a helical module having two long helices and two short helices that form a pocket lined with aromatic and aliphatic side chain residues. Consequently, the binding pocket in the small sub-domain side of the interface is very hydrophobic in its nature, whereas the large domain side of the interface contains residues that offer many possible hydrogen bond donor and acceptor sites, introducing therefore a polar region into the structure. This is illustrated in Fig. 2, where the surface is coloured brown to indicate lipophilicity and blue to indicate polarity.

The interface between the large and the small sub-domain is commonly subdivided into A, B and C sites in the sirtuins. This division is based on the binding of adenine, ribose and nicotineamide parts of the NAD^+ cofactor. The site for the adenine and ribose moieties has been well described in many studies (e.g. Sir2-Af1 [23, 38], Hst2 [36], Sir2-Af2 [25], SirT2 [22]), but the location of the nicotinamide moiety of the NAD^+ has been reported to be variable. Recently Avalos et al. [37] reported the binding of free nicotinamide co-crystallised at the C-site of Sir2Tm and Sir2-Af2, in a configuration where NAD^+ is proposed to be activated for catalysis. Catalysis is believed to take place when NAD^+ changes its conformation from a non-productive conformation [23, 25] to a productive conformation [24, 25]. In the productive conformation, the nicotinamide moiety is placed into the C-site.

The nicotinamide binding pocket is lined with "flexible loop" residues from 269 to 295 which include some of the most highly conserved residues in sirtuins.

This loop has been described to form a "frontwall"[37] of the C-site or a "ceiling"[23] on the pocket. The flexibility of the loop can be deduced from the different conformations that the loop can adopt (e.g. crystal structures of Sir2-Af1 [38], Sir2-Af2 [25], Hst2 [36]). Furthermore, it has been noted that when proteins are co-crystallised with NAD^+ complex, this results in an ordered loop orientation. For instance, Sir2-Af1 [38] and Hst2 [36] structures with ADP-ribose moiety of NAD^+ complexes may share the same ordered orientation, but without bound NAD^+, the loop possesses a more poorly defined or disordered orientation [36, 45, 46]. This also suggests that the presence of NAD^+ binding may trigger the assembly and disassembly of the front wall of the C pocket [37].

The C pocket is believed to be involved in a process known as nicotinamide exchange that results in enzyme inhibition [14, 25]. Whether the C pocket can be used as a binding site for developing new inhibitors is in focus at this study. The ordered loop orientation could be a precondition for the entrance of the nicotinamide or a small ligand into the pocket and to the inhibition of the enzyme. Therefore we used structures co-crystallised with NAD^+ or ADP-ribose for modelling the C pocket. Available SirT2 crystal structure (1J8F.pdb) does not include NAD^+, therefore we built the ordered conformation of the loop based on Sir2-Af1 (1M2G.pdb) and Hst2 (1Q1A.pdb) structures (Fig. 6b), which were fitted into SirT2 template structure (1J8F.pdb) [22]. Sir2-Af1 begins with an equal number of residues and high homology with SirT1 in this region, which meant that it was logical to adopt those coordinates for the beginning of the loop. The rest of the loop was then built by superimposing the structures of Hst2 and SirT2 and a 5-residue-long loop sought from the RCSB Protein Databank [26]. The local superimposition and more details are available in the supplementary data.

The rearrangement of the flexible loop has also been reported to be involved with rigid body rotation of the small domain relative to the large domain of the catalytic core [36]. Therefore NAD^+ binding to the Sir2-like proteins is believed to play both a catalytical and structural role in acetyl-lysine catalysis and binding. The rigid body rotation may be illustrated by overlaying the highly conserved Rossmann fold of Sir2-like proteins. This Rossmann fold superimposes well in all the homologs, whereas the smaller sub-domain exhibits variations in the location. The interface between these two sub-domains can be seen to work as a hinge. The movement of small sub-domain along this "hinge" may increase or decrease the 3-dimensions of the active site. However, more studies are needed to clarify the

importance of rigid body rotation in the deacetylation mechanism. In our model, the orientation was taken from a rigid body of the SirT2 (1J8F.pdb) structure.

Zinc binding module

The smaller sub-domain exhibits the most variation among the different homology structures. In case of SirT2 [22] it consists of a helical part and of a three-stranded antiparallel β-sheet in conjugation with the zinc-binding module. The zinc atom is tetrahedrally coordinated by the thiols of four Cys residues (371, 374, 395, 398 in SirT1). The four zinc-coordinating Cys residues are present in all classes of Sir2-like enzymes [22], but the remainder of the zinc-binding module is one of the least well-conserved parts of sirtuins present in the alignment. The most striking dissimilarity is the insertion of residues from D401 to A405 in SirT1. This insertion is adjacent to zinc finger coordination. None of the compared homology structures have insertion at this site, and the generation of the site was not an option due to restriction in the three dimensional space of this site. We wanted to maintain the zinc finger coordination and therefore the backbone of the corresponding SirT2 template structure was taken directly at this site, even though it differs in its amino acid sequence (Figs. 5, 6a). This area is rather far away from the binding site and thus some uncertainty in that area in the model of SirT1 was anticipated to have little impact on the active site.

The binding mode

A series of indoles [19, 20] have been docked in the C-pocket of our SirT1 homology model (See Table 1). In order to illustrate the orientation of the docked compounds, the binding mode of the S-enantiomer of the most potent inhibitor (EX527), compound 1(S) in Table 1, is depicted in Figs. 2 and 7. According to docking, compound 1(S) binds to these available H-bond acceptor/donor sites, with all its donor and acceptor atoms bound. The aromatic and aliphatic parts of the structure fit well into the hydrophobic environment.

The complex is characterized by an extensive H-bond network. As shown in Fig. 7, the amidic NH_2 of the inhibitor donates H-bonds to the carboxyl group of D348. In addition, the carbonyl oxygen forms H-bridges to both of the backbone NHs of D348 and I347. The NH portion of the indole ring serves as an H-bond donor to the backbone carbonyl oxygen of Q345. Next to this H-bond network, the binding pocket is built up by the flexible loop in its ordered conformation triggered by NAD^+ binding [23, 36, 38]. In our

model, the F273 residue is orientated into the cavity and restricts the size of the ligand. On the opposite site of the H-bond network, the chlorine of the aromatic ring of the inhibitor comes into contact with H363 and P293. Next to this, the binding pocket consists of several hydrophobic residues, I279, M296, F309, F312, I316 and P409. For reasons of clarity, those residues have been omitted from the figure.

Interestingly, the binding site for compound 1(S) occupies the same area as that reported for the ribose and nicotinamide moieties of NAD^+ [22, 23, 37]. Finnin et al. [22] have also reported that mutations on SirT2 residues H187, Q167, N168 and D170 (H363, Q345, N347 and D348 correspondingly on SirT1) disrupt the histone deacetylase activity. Similar mutational studies have also been reported with other homologs [23, 38, 47]. These findings support the binding mode of compound 1(S) proposed in this study.
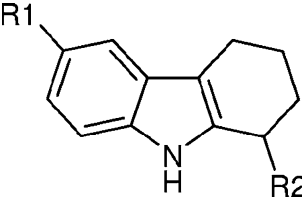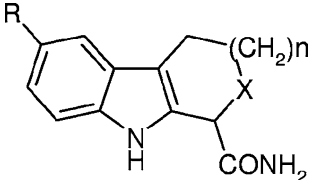
We also studied other docking solutions of compound 1 to find alternatives for the preceding binding mode. Among the 10 independent docking runs, we found that all the docking solutions of compound 1 are in the same plane as the best-docked conformation described above. Sorting of these solutions into a logical order was done according to GoldScore scoring function. It became clear that the decreasing number of the H-bonds was relatively to the lower scoring. However no other H-bonds than above mentioned were detected between the ligand and the enzyme. This made it reasonable to choose the presented binding mode as the bioactive conformation.
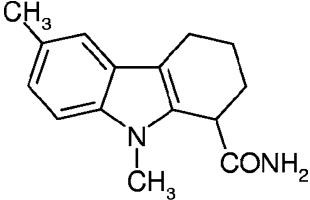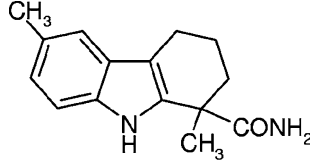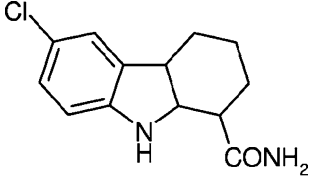
The pharmacophore

We docked the S and R enantiomers of the structures presented in Table 1 at the binding site and the best-docked configuration of each compound was selected according to the GoldScore fitness function. The docked compounds showed only little variation in the orientation of the tricyclic moiety. In most cases the ring system is in the same plane as described with compound 1(S). Among the best-docked structures only the R enantiomer of the compound 8 (inactive) differed significantly from this orientation, building up H-bonds between NH group of indole and D348. Overall the conformations of the docked compounds showed to have H-bonds only with the residues Q345, I347 and D348. It was also striking that the less potent or non-potent ligands showed only some of the possible H-bonds of the compound 1(S).

The docked compounds were subdivided according to their reported biological data into three classes: high potent, potent and inactive (see Table 2 for details).

Table 1 Indole derivates and inhibition of SirT1[a]

Compound	R1	R2	IC$_{50}$ (μM)[a]	Compound	R	n	X	IC$_{50}$ (μM)
1	Cl	CONH$_2$	0.098	20	Cl	1	NH	18.0
2	CH$_3$	CONH$_2$	0.205	21	Br	1	NH	>100
3	H	CONH$_2$	1.47	30	Cl	0	CH$_2$	0.409
5	Cl	CO$_2$Et	>100	31	H	0	CH$_2$	2.67
8	Cl	CO$_2$H	>100	35	Cl	2	CH$_2$	0.124
12	OH	CONH$_2$	15.0					
15	Cl	CONHOH	77.6	Compound	R	n	X	IC$_{50}$ (μM)[a]
36	Cl	CONHEt	>100	40	–	–	–	>100
37	Cl	CONEt$_2$	>100	41	–	–	–	13.0
38	Cl	CONHCH$_2$CONH$_2$	>100	13	–	–	–	>100
39	CH$_3$	CH$_2$NH$_2$	>100					

40

41

13

[a] The activity data for rasemic compounds is reported according to Napper et al. [19]

No correlation between the scoring function and the biological activity (IC$_{50}$-value) was detected. After that, as a post-docking filter, we examined several combinations of the previously described features of the binding mode with the aim to find the simplest combination that would describe the active compounds as hits and the inactive molecules as non-hits. It transpired that the features of the H-bond network had to be fulfilled. In other words, the pharmacophore consists of H-bonds to the carboxyl group of D348, to the backbone carbonyl oxygen of Q345 and one H-bond to the NH backbone of either D348 or I347. It also seems that all compounds defined as hits occupy the space close to H363. Whether this spatial orientation is a precondition of the pharmacophore cannot be determined with the available set of inhibitors.

Because the pharmacophore was designed for virtual screening of databases, the screening protocol had to be simple and comprehensive. Only the best docked configuration of each compound was selected according to the GoldScore fitness function. The activity/inactivity of the docked structures was then evaluated using the post-docking filter, which was based on

pharmacophore considerations. The performance of the model is presented in Table 3. The compounds which fulfilled the requirements of the pharmacophore are denoted as hits and the compounds which did not fulfil the requirements are denoted as non-hits.

The pharmacophore is able to screen out all the non-potent compounds of the dataset by only taking the best docking conformation into account. For instance, compound 8, which differs from compound 1 having a carboxyl group in R2 position instead of an amide group is de-protonated and therefore it does not fulfil the requirement of the pharmacophore. There is no H-bond between the ligand and D348. Compound 5 is also inactive as evidence that a hydrogen bonding donor (to D348) is needed at this position in the cyclohexene ring. Compound 13 has a saturated cyclohexene ring which may have altogether six different stereoisomers. We could not find any that fulfil the pharmacophore. Obviously the double bond is important for the steric orientation of the amide group.

Eleven potent compounds were screened and eight of them were identified as hits. Three were screened out as non-hits (S2, S3, R35), because only the best-
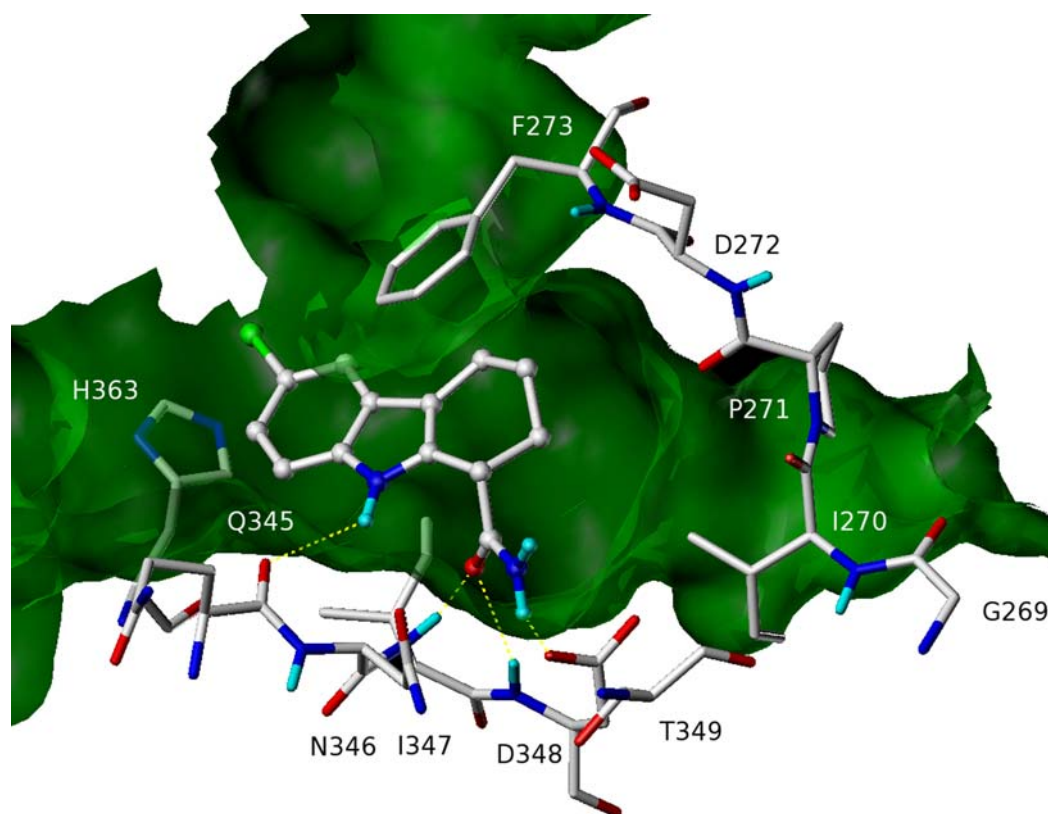
Fig. 7 The binding mode of EX527 (compound 1(S))

docked conformation was taken into account. When the other docking conformations were taken into account also the conformation that fulfilled the requirements of the pharmacophore was found. For instance, the fifth best scoring conformation of the compound 2 fulfilled the requirements. However, it is not uncommon that some potent compounds are screened out, when using simple protocols.

Table 2 Classification of tested compound according to their biological activity[a]

	IC$_{50}$ range	Members
High potent	IC$_{50}$ ≤ 1 μm	S1, S2, S30, S35
Potent	1 μm < IC$_{50}$ ≤ 100 μm	S3, S12, S20, S31, S41, S15, R35
Inactive	IC$_{50}$ > 100 μm	R1, R2, 5, 8, 13, 21, 36, 37, 38, 39, 40

[a] If no configuration is given, the racemate is inactive. Only the absolute configuration of compound 35 and the IC$_{50}$ values of the pure enantiomers have been reported. The S-enantiomer is 350 times more potent than the R-enantiomer. Since the core structure of all compounds are basically the same we conclude, that in all cases the more active configuration has the S configuration at the substituted alkyl ring

The molecular interaction fields

Molecular interaction fields in the active site region were calculated using the GRID software package [33]. The fields of all multi-atom-probes and all single-atom-probes, except for metallic cations were calculated. Residues involved in creating the binding pocket contain some of the most extensively conserved residues in the sirtuins G261–S265, G269–R274, T344–D348 and H363 (see the highlighted areas in Fig. 5). Strong hydrophobic interactions were noted at the small subdomain site of the interface using aliphatic or aromatic groups, methylprobe and dryprobe. Strong polar

Table 3 Selection based of the pharmacophore filter[a]

	Hit	Non-Hit
High potent	S1, S30, S35	S2
Potent	S12, S20, S41, S15, S31	S3, R35
Inactive		All inactive compounds

[a] The model only takes account that a compound is a hit or a non hit, the classification to potent and high potent compounds is according to biological data

interactions relative to the water–probe interaction were observed mainly at the large sub-domain side of the binding cavity.

In particular, the area around residues T344 to D348, also described as a "floor" of the binding pocket [23] appeared to have possible interaction sites for ligand binding (see Fig. 7). Near to D348 carboxyl, cationic amine probes showed considerably strong interaction energies. Neutral amines or amines with lone pairs showed lower interaction energies at that site. Strong interactions with H-bond donating nitrogen probes were also provided on the C-site front wall where the backbone carbonyls of residues P271 and D272 in the flexible loop may act as H-bond acceptors.

In conclusion, H-bond donating groups and the basicity of the ligands are proposed as being important factors affecting binding in the C-site. The strong interactions detected with the amine probes are consistent with the proposed binding mode of compound 1(S) found in this study (Fig. 7). In addition to the carboxyl group–amine interaction, proton acceptors have the possibility to undergo hydrogen bonding to the backbone NHs of D348 or I347. Therefore it is understandable why certain ligands which possess both donor and acceptor functionalities such as the amide group in nicotineamide or in compound 1(S) can bind to the active site of SirT1.

Conclusion

A comparison model of the human sirtuin SirT1 has been constructed using human sirtuin SirT2 as the primary template. MD simulation for 400 ps at 300 K was run to refine the orientations of the side-chains. The generated model shares a similar backbone structure with SirT2, having all of its β-sheets and most of the α-helices assigned directly from the SirT2 structure. The most significant difference when compared to SirT2 is the flexible loop conformation modeled to have an ordered conformation according to homology structures co-crystallised with the ADP-ribose moiety of NAD^+.

A number of indoles [19, 20] have been docked at the active site of our SirT1 homology model and a binding mode based on an extensive H-bond network was found. EX527 was found to bind in the very extensively conserved region of residues from T344 to D348. The pharmacophore consists of H-bonds to the carboxyl group of D348, to the backbone carbonyl oxygen of Q345 and one H-bond to the NH backbone

of either D348 or I347. Also the space in the vicinity of H363 may be considered as being important.

References

1. Ekwall K (2005) Trends Genet 21:608
2. Guarente L (2000) Genes Dev 14:1021
3. Sauve AA, Celic I, Avalos J, Deng H, Boeke JD, Schramm VL (2001) Biochemistry 40:15456
4. North BJ, Marshall BL, Borra MT, Denu JM, Verdin E (2003) Mol Cell 11:437
5. Blander G, Guarente L (2004) Annu Rev Biochem 73:417
6. Porcu M, Chiarugi A (2005) Trends Pharmacol Sci 26:94
7. Luo J, Nikolaev AY, Imai S, Chen D, Su F, Shiloh A, Guarente L, Gu W (2001) Cell 107:137
8. Brunet A, Sweeney LB, Sturgill JF, Chua KF, Greer PL, Lin Y, Tran H, Ross SE, Mostoslavsky R, Cohen HY, Hu LS, Cheng HL, Jedrychowski MP, Gygi SP, Sinclair DA, Alt FW, Greenberg ME (2004) Science 303:2011
9. Nemoto S, Fergusson MM, Finkel T (2005) J Biol Chem 280:16456
10. Rodgers JT, Lerin C, Haas W, Gygi SP, Spiegelman BM, Puigserver P (2005) Nature 434:113
11. Howitz KT, Bitterman KJ, Cohen HY, Lamming DW, Lavu S, Wood JG, Zipkin RE, Chung P, Kisielewski A, Zhang LL, Scherer B, Sinclair DA (2003) Nature 425:191
12. Kaeberlein M, McDonagh T, Heltweg B, Hixon J, Westman EA, Caldwell SD, Napper A, Curtis R, DiStefano PS, Fields S, Bedalov A, Kennedy BK (2005) J Biol Chem 280:17038
13. Borra MT, Smith BC, Denu JM (2005) J Biol Chem 280:17187
14. Bitterman KJ, Anderson RM, Cohen HY, Latorre-Esteves M, Sinclair DA (2002) J Biol Chem 277:45099
15. Grozinger CM, Chao ED, Blackwell HE, Moazed D, Schreiber SL (2001) J Biol Chem 276:38837
16. Marcotte PA, Richardson PR, Guo J, Barrett LW, Xu N, Gunasekera A, Glaser KB (2004) Anal Biochem 332:90
17. Bedalov A, Gatbonton T, Irvine WP, Gottschling DE, Simon JA (2001) Proc Natl Acad Sci USA 98:15113
18. Mai A, Massa S, Lavu S, Pezzi R, Simeoni S, Ragno R, Mariotti FR, Chiani F, Camilloni G, Sinclair DA (2005) J Med Chem 48:7789
19. Napper AD, Hixon J, McDonagh T, Keavey K, Pons JF, Barker J, Yau WT, Amouzegh P, Flegg A, Hamelin E, Thomas RJ, Kates M, Jones S, Navia MA, Saunders JO, DiStefano PS, Curtis R (2005) J Med Chem 48:8045
20. Solomon JM, Pasupuleti R, Xu L, McDonagh T, Curtis R, Distefano PS, Huber LJ (2006) Mol Cell Biol 26:28
21. Frye RA (2000) Biochem Biophys Res Commun 273:793
22. Finnin MS, Donigian JR, Pavletich NP (2001) Nat Struct Biol 8:621
23. Min J, Landry J, Sternglanz R, Xu RM (2001) Cell 105:269
24. Zhao K, Harshaw R, Chai X, Marmorstein R (2004) Proc Natl Acad Sci USA 101:8563
25. Avalos JL, Boeke JD, Wolberger C (2004) Mol Cell 13:639
26. Berman HM, Battistuz T, Bhat TN, Bluhm WF, Bourne PE, Burkhardt K, Feng Z, Gilliland GL, Iype L, Jain S, Fagan P, Marvin J, Padilla D, Ravichandran V, Schneider B, Thanki N, Weissig H, Westbrook JD, Zardecki C (2002) Acta Crystallogr D Biol Crystallogr 58:899
27. Protein-protein BLAST (blastp), National Center for Biotechnology Information, Bethesda, U.S. (http://www.ncbi.nlm.nih.gov/BLAST/Blast.cgi)

28. SAS, European Bioinformatics Institute, Cambridge, U.K. (http://www.ebi.ac.uk/thornton-srv/databases/sas/)

29. InsightII v 2000, Accelrys, Inc., San Diego, CA (http://wwwaccelryscom/)

30. GROMACS, v. 3.1.4, Department of Biophysical Chemistry, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands (http://www.gromacs.org/)

31. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ (2005) J Comput Chem 26:1701

32. Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM (1996) J Biomol NMR 8:477

33. GRID v. 1.2.2, Molecular Discovery Ltd., U.K. (http://www.moldiscovery.com/)

34. GOLD v. 3.0, Cambridge Crystallographic Data Centre, Cambridge, U.K. (http://www.ccdc.cam.ac.uk)

35. SYBYL v. 7.1, Tripos Inc., St. Louis, MO (www.tripos.com)

36. Zhao K, Chai X, Marmorstein R (2003) Structure 11:1403

37. Avalos JL, Bever KM, Wolberger C (2005) Mol Cell 17:855

38. Chang JH, Kim HC, Hwang KY, Lee JW, Jackson SP, Bell SD, Cho Y (2002) J Biol Chem 277:34489

39. Zhao K, Chai X, Marmorstein R (2004) J Mol Biol 337:731

40. GeneDoc v. 2.6.002, Nicholas KB, Nicholas HB (http://www.psc.edu/biomed/genedock)

41. Essman U, Perrera L, Berkowitz ML, Darden T, Lee H, Pedersen LG (1995) J Chem Phys 103:8577

42. Shan Y, Klepeis JL, Eastwood MP, Dror RO, Shaw DE (2005) J Chem Phys 122:54101

43. Berendsen HJC, Postma JPM, Van Gusteren WF, DiNola A, Haak JR (1984) J Chem Phys 81:3684

44. Tervo AJ, Kyrylenko S, Niskanen P, Salminen A, Leppanen J, Nyronen TH, Jarvinen T, Poso A (2004) J Med Chem 47:6292

45. Avalos JL, Celic I, Muhammad S, Cosgrove MS, Boeke JD, Wolberger C (2002) Mol Cell 10:523

46. Finnin MS, Donigian JR, Cohen A, Richon VM, Rifkind RA, Marks PA, Breslow R, Pavletich NP (1999) Nature 401:188

47. Avalos JL, Bever KM, Wolberger C (2005) Mol Cell 17:855