J Comput Aided Mol Des (2006) 20:621–628
DOI 10.1007/s10822-006-9062-2

ORIGINAL PAPER

Molecular modeling and bioinformatical analysis of the antibacterial target enzyme MurA from a drug design perspective

Christian D. Klein · Anke Bachelier

Received: 21 June 2006 / Accepted: 28 July 2006 / Published online: 24 November 2006

Abstract The enzyme MurA (UDP-*N*-acetylglucosamine enolpyruvyl transferase) catalyzes the first cytoplasmatic step in the synthesis of murein precursors. This function is of vital relevance for bacteria, and the enzyme therefore represents an important target protein for the development of novel antibacterial compounds. Several X-ray structures of liganded and un-liganded MurA have been published, which may be used for rational drug design. MurA, however, contains a highly flexible surface loop, which is involved in substrate and inhibitor binding. In the available X-ray structures, the conformation of this surface loop varies, depending on the presence or absence of ligands or substrate and probably also on the crystal packing. The uncertainty of the low-energy, or ''resting state'' conformation of this surface loop hampers the application of rational drug design to this class of enzymes. We have therefore performed an extensive molecular dynamics study of the enzyme in order to identify one or several low-energy conformers. The results indicate that, at least in some of the X-ray structures, the conformation of the flexible surface loop is influenced by crystallographic contacts. Furthermore, three partially helical foldamers of the surface loop are identified which may resemble the resting states of the enzyme or intermediate states that are ''traversed'' during the substrate binding process. Another, very important aspect for the development of novel antibacterial compounds is the inter- and intra-species variability of the target structure. We present a comparison of MurA sequences from 163 organisms which were analyzed under the aspects of enzyme mechanism, structure and drug design. The results allow us to identify the most promising binding sites for inhibitor interaction, which are present in MurA enzymes of most species and are expected to be insusceptible to resistance-inducing mutations.

Keywords Antibiotics · Drug design · Enzyme inhibition · Molecular dynamics simulation · Multiple sequence alignment · MurA · EC 2.5.1.7

Abbreviations

MurA	UDP-*N*-acetylglucosamine enolpyruvyl transferase
UNAG	UDP-*N*-acetylglucosamine
PEP	phosphoenolpyruvate
RMSD	root-mean-square deviation
MD	molecular dynamics

Introduction

The UDP-*N*-acetylglucosamine enolpyruvyl transferase (MurA, EC 2.5.1.7) catalyzes the first step in the cytoplasmatic biosynthesis of murein precursors. This reaction is of vital importance for bacteria [1]. Because there is no analogous metabolic pathway or enzyme in eukaryotes, MurA represents a favorable point of attack for novel antibacterial drugs. This strategy is under active development in a number of research groups.

C. D. Klein (✉) · A. Bachelier
Pharmazeutische und Medizinische Chemie, Saarland University FR 8.2, P.O. Box 15 11 50, Am Stadtwald, D-66041 Saarbrücken, Germany
e-mail: cdpk@mx.uni-saarland.de

MurA consists of a single amino acid chain that forms two globular domains with similar folds (cf. Fig. 1) [2, 3]. The active site lies between the two domains and is closed upon the binding of the substrates UDP-*N*-acetylglucosamine (UNAG) and phosphoenolpyruvate (PEP). This involves not only the movement of the two domains towards each other, but also the closure of the lid-like surface loop containing the catalytically important residue Cys115. The surface loop is terminated by two proline ''hinges'' that are present in nearly all known MurA sequences. In *E. coli*, the two prolines are in positions 112 and 121, respectively. The numbering from the *E. coli* sequence will be used to identify the key residues throughout this article, although the analogous residues may have a slightly different position in the MurA of other species. In this work, the surface loop will be termed ''Cys115 loop''.

MurA is selectively inhibited by the PEP-mimicking antibiotic fosfomycin, which forms a covalent bond to the sulfhydryl group of the catalytically relevant Cys115 residue [2, 4]. The binding of fosfomycin is greatly facilitated in presence of UNAG, because UNAG promotes the closing of a flexible surface loop (which also contains the fosfomycin-binding cysteine residue) to form the catalytically competent ''closed'' conformer of the enzyme. The antibiotic activity of fosfomycin therefore depends on the presence of the Cys115 residue, which, however, can be replaced by aspartate without loss of catalytic activity. Therefore, resistance develops quickly, and some pathogens (like *M. tuberculosis*), which natively have an aspartate in position 115, are resistant towards fosfomycin.

Several X-ray structures of MurA in closed and open states have been published over the last few years (cf. Table 1). The value of these structures for rational drug design, however, is limited because of the innate flexibility of the Cys115 loop. This loop constitutes a substantial part of the active site of the enzyme. Therefore, any structure-based, theoretical drug-design approach will require that the loop is positioned in a low-energy conformation.

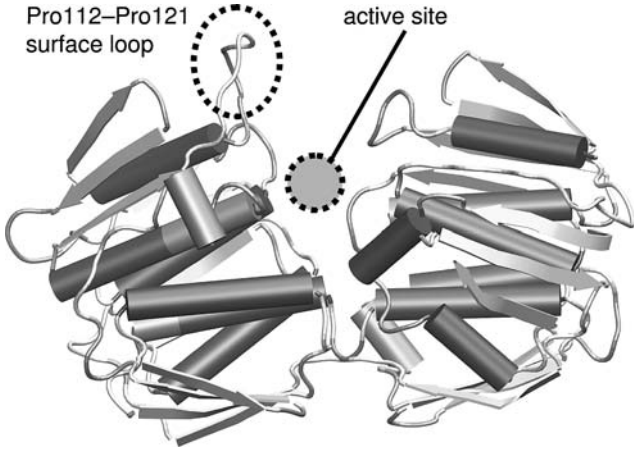

Fig. 1 MurA in the ''open'' conformation. Upon substrate binding, the two domains approach each other and the Cys115 surface loop closes the active site. PDB code 1NAW [3]

One tempting suggestion for MurA inhibitor development has been to design ligands that bind to the open or semi-closed conformer of the Cys115 loop, thereby stabilizing the enzyme in the resting state [5]. For the computer-assisted identification or design of such compounds, however, it will be of supreme importance to identify one or more low-energy conformers which could then be used as a structural basis for virtual screening or other computational methods.

Retrospective analysis of available X-ray structures

The purpose of this part of the work was to study the influence of contacts in the crystal lattices on the conformation of the Cys115 loop. Packing effects have already been discussed for some structures, e.g. in [6], but there exists no comprehensive overview for the available MurA structures. In order to assess the value of the published X-ray structures for the rational design of new ligands, a careful analysis of the highly flexible Cys115 loop must be performed.

The published X-ray structures of MurA enzymes are summarized in Table 1. The structures were either obtained for the *E. coli* enzyme or for the closely related MurA from *Enterobacter cloacae*. Images of the crystal contact regions are shown in the supporting information. It is evident from Table 1 that the loop closure depends on the presence of ligands, such as UNAG/fosfomycin or reaction intermediates. In the closed state, very few crystal contacts—if any—are observed for the Cys115 loop. A different picture emerges for the un-liganded forms of MurA and for complexes with compounds that do not bind in a substrate-like position. In those cases, the loop remains open or semi-closed and usually makes a number of contacts to adjacent protein molecules. This is especially true for the structures 1NAW (cf. Fig. 1), 1EJD, 1DLG, and 1YBG, where the Cys115 loop protrudes radially away from the protein globule and is partially surrounded by residues of the adjacent monomer. In these four cases, the contacts to Lys131, Val157 and Asp159 of the neighbor protein are most prominent.

In conclusion, there is clear evidence that conformation of the Cys115 loop in the un-liganded MurA X-ray structures depends at least to some degree on

Table 1 Published X-ray structures of MurA enzymes

PDB-code	Loop state	Species	Ligand	Residue numbers of adjacent monomers[a]	Reso-lution	Mutation	Ref.
1UAE	Closed	*E. coli*	UNAG and fosfomycin	None	1.80 Å	None	[2]
1A2N	Closed	*E. coli*	Tetrahedral reaction intermediate	D159 and K160 within 6 A	2.80 Å	Cys115Ala	[7]
1Q3G	Closed	*E. cloacae*	Tetrahedral reaction intermediate	332, 333, 336, 337 **Reaction intermediate**	2.65 Å	Asp305Ala	[8]
1RYW	Closed	*E. cloacae*	Reaction product	330–340 **E337, reaction product**	2.30 Å	Cys115Ser	[9]
1NAW	Open	*E. cloacae*	None	130–133, 155–160 **K131, V157, D159**	2.00 Å	None	[3]
1DLG	Open	*E. cloacae*	None	127, 128, 131, 115–169 **K131, V157**	1.90 Å	Cys115Ser	[10]
1EJC	Open	*E. cloacae*	None	37–40, 199–202, 220–225 **A200, G223**	1.80 Å	None	[6]
1EJD	Open	*E. cloacae*	None	124, 127, 128, 131, 155–160 **K131, V157, D159** from one, **E348** from another neighbor	1.55 Å	None	[6]
1EYN	Open	*E. cloacae*	8-Anilino-1-naphthalene-sulfonate (ANS)	38–40, 65–67, 71–72, ANS **ANS fluorescence probe**	1.70 Å	None	[5]
1YBG	Open	*E. cloacae*	Inhibitor T6361	123–125, 127–131, 155–160 **V157, D159**	2.60 Å	None	[11]

[a] Printed in bold: residues of adjacent proteins or name of the ligand molecule within 3 A of the surface loop

crystal contacts and may therefore not be a valid starting point for computer-based drug design.

Molecular dynamics simulations: strategy and setup

In order to explore the conformational space that is available to the Cys115 loop, to study the movement of the two domains and to identify possible low-energy conformers, extensive molecular dynamics simulations starting from one of the *E. cloacae* open-loop structures (1NAW) were carried out. The *E. cloacae* MurA was chosen because a variety of open and closed three-dimensional structures—which can be used to cross-check the results of the simulations—have been published for this enzyme. In any case, it must be noted that the *E. coli* and *E. cloacae* enzymes are extremely similar (93.6% sequence identity) and that their plasticity and conformational behavior can therefore be assumed to be analogous, if not identical.

The protocol of the simulation is given in Table 2. The ''Generalized Born'' implicit-solvent solvation model was used for all simulations in order to avoid the retarding effect of explicit solvent molecules. Therefore, the time scale has only limited physical significance. Compared to explicit-solvent interactions, a much better sampling of conformational space can be achieved by using high-level implicit-solvent models such as the Generalized Born approach.

After the initial preparatory work (stage 1), the simulation system was heated to 300 K and 2 ns of unrestrained dynamics was performed (stage 2). The resulting system was then subjected to restraints for the peptide backbone atoms, except for the Cys115 loop, and heated to 450 K (stage 3, 2 ns simulation) to effectively explore the conformational space available to this loop. The restraints for the backbone atoms of the two globular domains were necessary in order to avoid the complete unfolding of the protein. From the resulting trajectory, 200 frames were extracted as representative snapshots of the dynamical profile of the protein (stage 4). These snapshots were then minimized (stage 5) to yield an ensemble of structurally diverse, local-minimum conformers to be analyzed in stage 6.

Unrestrained molecular dynamics simulations

The potential energy plot given in Fig. 2 demonstrates that the energy is well-equilibrated after about 2×10^6 steps. Concerning the dynamic properties of the system, the RMSD plot (documented in the supporting

Table 2 The MD simulation protocol. Further details of the simulations are provided as supporting information

Stage	Action	Key conditions
1	Removal of solvent and ligands; addition of hydrogens; minimization	–
2	2 ns MD, initial backbone restraints suspended after 100 ps	300 K, step size 1.5 fs
3	2 ns MD starting from final state of stage 2, backbone restrained except Cys115 loop	450 K, step size 1 fs
4	Frames from stage 3 extracted at 10 ps-intervals	–
5	Minimization of frames from stage 4	400 steps steepest descent, 1,600 steps conjugate gradient
6	Analysis of minimized frames	–

information) indicates a second, faint equilibration phase between 1×10^6 and 2×10^6 steps. Figure 2 also displays the variation of the interdomain distance between. It is very clear that the two domains approach each other upon relaxation of the restraints. The movement is initially very fast. The second phase of the domain movement involves reorganization of the side chains in the interdomain cleft and therefore proceeds more slowly. Eventually, the interdomain distance levels out on a plateau that is about 3 Å below the initial state (cf. Fig. 3).

At the beginning of the simulation (minimized crystal structure), the Cys115 loop is in a practically unfolded conformation. As soon as the restraints are released, it begins to fold and settles into a relatively stable, partially helix-like structure (cf. Fig. 4). Notably, the short helical region is terminated by two glycines on the N-terminal and one glycine on the C-terminal end. These glycines are strictly conserved in all currently known MurA sequences (see below for a more detailed discussion of sequence homologies).

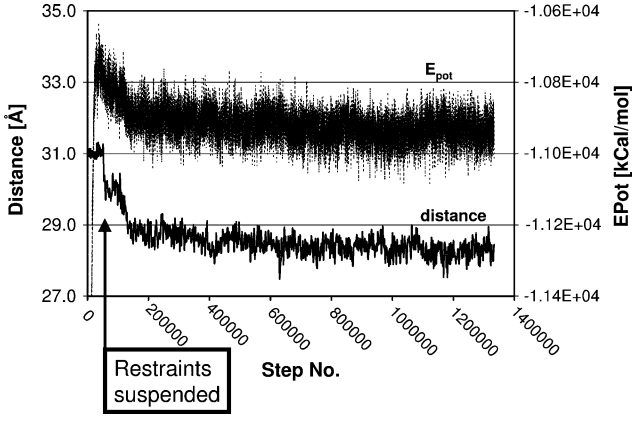

Fig. 2 Plot of potential energy and interdomain distance for the first, unrestrained MD simulation (stage 2). The interdomain distance (lower graph) was measured between the centers of mass of the two domains

High-temperature molecular dynamics simulations

A molecular dynamics simulation at higher temperature (450 K) were performed in order to obtain a more rigorous sampling of the conformational space that is available to the Cys115 loop (stage 4). The simulation started with the endpoint of the previous simulation. All backbone atoms not belonging to the Cys115 loop were restrained to their initial positions to prevent the complete unfolding of the protein. The Cys115 loop attains considerable flexibility in the high-temperature simulations, as can be seen from the superimposed structures in Fig. 5 and the RMSD plot (cf. supporting information). Two hundred frames from the high-temperature simulation were minimized (stage 5) and the resulting low-energy structures were subjected to a cluster analysis in order to obtain a representative overview of the most relevant conformers (stage 6) (Fig. 6).

Cluster analysis

The cluster analysis yielded 27 clusters, which mostly consisted of no more than three conformers. Only the three significantly populated clusters containing 85, 42, and 45 members, respectively, will be analyzed in detail. The average potential energies of the cluster members were –15678 ± 43, –15691 ± 33, and –15689 ± 30 kcal/mol (values ± standard deviation). The average potential energy of the first cluster (85 conformers) is somewhat lower than that of the others, but the significance of this difference is doubtful, given the approximative nature of force-field-based simulations. Representative members from each of the clusters are shown in Fig. 7. These conformers obviously differ significantly from each other. They also differ from the starting conformation where the loop was nearly unfolded. Thus, we can conclude that there is a clear tendency to adopt a (partially) folded conformation, but we can not decide with certainty which of the resulting clusters resembles a

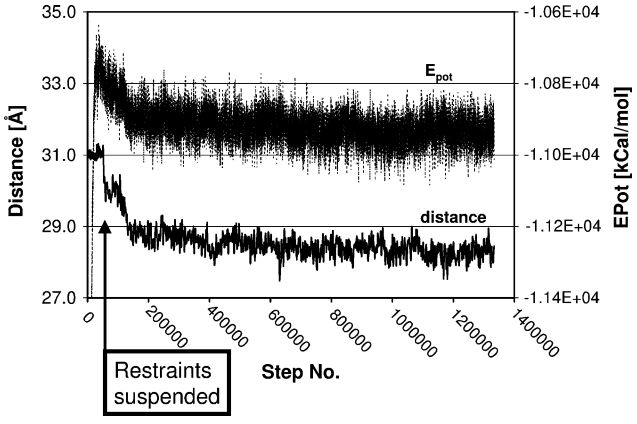

Fig. 3 Synoptic representation of the dynamic behavior of MurA during the unrestrained molecular dynamics simulation. The small arrow in the first picture indicates the Cys115 surface loop

hypothetical *single* resting state of the enzyme. In the context of docking and other drug design efforts, this result means that theoretical methods (docking, de-novo-design) aimed at single conformers do not seem promising.

Sequence comparisons

At the beginning of 2006, 163 sequences of MurA enzymes from various microorganisms were available on the Swiss-Prot web server (UniProt). The homology between those enzymes varies to a large extent, with some MurA enzymes being extremely similar to each other (nearly 95% identity for *E. coli* and *E. cloacae*), although the average homology amounts to about 30%. This large number of sequences, all of which ultimately provide a functional enzyme, allows us to identify those residues with the largest functional or structural relevance, because these residues should be highly conserved.

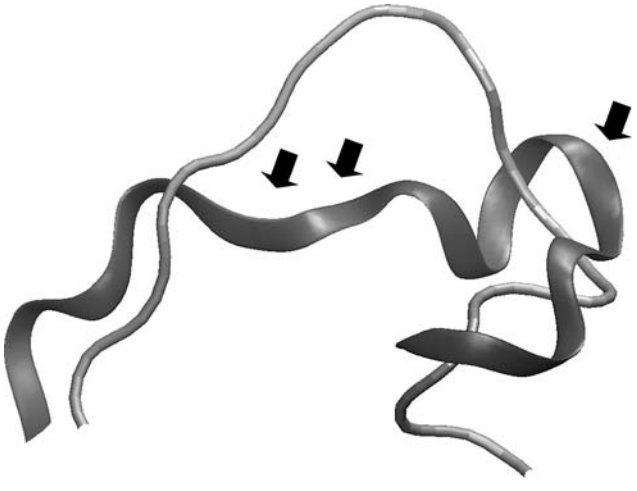

Fig. 4 The Cys115 surface loop at the beginning (thin tube) and at the end (ribbon) of the unrestrained MD simulation. The positions of the strictly conserved Glycines 113, 114 and 118 are highlighted by arrows

The multiple alignment of the 163 sequences showed that 20 residues are completely conserved. Among these are eight glycine residues, which mostly act as helix-terminators. Gly398 terminates a helix near the active site and may therefore be of some relevance for the development of novel inhibitors. The graphical overview given in Fig. 8 shows a relatively large degree of conservation not only in the active site, but also at certain peripheral regions, where glycines and (more rarely) prolines serve as building blocks with specialized geometrical properties. There are also three conserved glycines (113, 114, 118) in the Cys115 loop, where they obviously serve as flexible hinge groups, allowing the surface loop to change its folding state during the catalytic cycle. Arg120, near the end of the surface loop, is also completely conserved, whereas the terminator prolines (112, 121) are less strictly conserved. Arg120 plays a pivotal role in the transition from the open to closed conformation by stabilizing the closed conformation though interaction with the phosphonate group of substrate UNAG [12]. Substitution to the basic amino acid lysine yields the enzyme unable for the conformational change into the closed state which underlines the importance of a conservation of this important residue among different bacterial strains. Cys115, the target of fosfomycin, is replaced by aspartate in about 20% of the MurA sequences. The other residues within the Cys115 loop are even less conserved, with, for example, Ala119 being replaced by various other residues in 60% of the sequences.

It is not surprising to find the highest degree of conservation at the bottom of the interdomain cleft, the binding region of the substrates UNAG and PEP (or the PEP-analog fosfomycin). Numerous residues with highly functionalized side chains are located in this area: Lys22, Asn23, Asp49, Asp231, Asp305, Arg331, Arg371, and Arg397 are completely conserved.

Conclusions

We have shown that the Cys115 surface loop has two distinctive properties that reduce its value as target structure for rational drug design.

First, it is a highly flexible region. Compounds that bind to and ''freeze'' this part of the enzyme will have

Residue #	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131
Cons. Seq.	?	V	S	L	P	G	G	C	A	I	G	A	R	P	V	D	L	H	I	K	G	L	E	?
Abundance (%)	----	84.7	66.9	84.7	95.7	100.0	100.0	84.7	70.6	95.7	100.0	39.9	100.0	86.5	72.4	87.1	53.4	93.9	54.6	30.7	72.4	87.1	47.9	----
E. coli	Q	V	S	L	P	G	G	C	T	I	G	A	R	P	V	D	L	H	I	S	G	L	E	Q
E. cloacae	Q	V	S	L	P	G	G	C	A	I	G	A	R	P	V	D	L	H	I	F	G	L	E	K

Fig. 5 Conservation of amino acids in and near the Cys115 surface loop of MurA enzymes. The table cells are colored according to degree of conservation, from 100 % (red) to less than 50 % (white). The residue numbering is for *E. coli* MurA. ''Cons. Seq.'' is the most frequent amino acid in that particular position of the 163 MurA enzymes. The sequences of the *E. coli* and *E. cloacae* enzymes are given at the bottom

to provide free energy to account for the prominent loss of entropic freedom of the Cys115 loop. Such compounds will therefore have a considerably lower affinity than compounds that bind to a less flexible part of the enzyme. From a drug-design perspective, the value of the X-ray structures with an open-loop conformation is limited because crystal contacts influence the geometry of the loop to a variable—and in many cases considerable—extent. It is also doubtful whether the interdomain cleft of the native enzyme in the resting state opens as wide as in the X-ray structures of un-liganded MurA.

Second, it is important to consider the degree of conservation of the residues in the Cys115 loop. An ideal target region should be highly conserved—this is a paramount prerequisite to obtain a wide-spectrum drug that cannot easily be rendered ineffective by resistance-inducing mutations. In the Cys115 loop, the degree of conservation of residues that are suitable interaction partners is very limited. The only completely conserved amino acid with relevant potential for interactions with a drug molecule is Arg120. Besides that, the loop contains three absolutely conserved glycine residues, but this amino acid has only a limited potential for making exergonic interactions. The other residues, including Cys115 which could be an important binding site for the design of non-covalent inhibitors, are more or less variable. Most of those residues are also not rich in chemical functionality and therefore not very amenable to site-specific drug design.

In contrast, most residues at the bottom of the interdomain cleft are highly conserved and very feature-rich. Furthermore, this region is much less flexible and there exist various X-ray structures with bound substrate analogs that can be used as starting points for structure-based drug design. We therefore conclude that a drug development effort targeted on this region is considerably more promising than to aim at the Cys115 surface loop.

The bottom of the interdomain cleft contains numerous acidic or basic residues. With respect to potential inhibitors, this would entail a demand for charged or at least highly polar functional groups as in fosfomycin. The presence of such groups, in turn, may lead to poor pharmacokinetic properties of the ligand. Therefore, a particular challenge with MurA inhibitors that bind to the interdomain cleft will be to find the right balance between target affinity and pharmacokinetics.

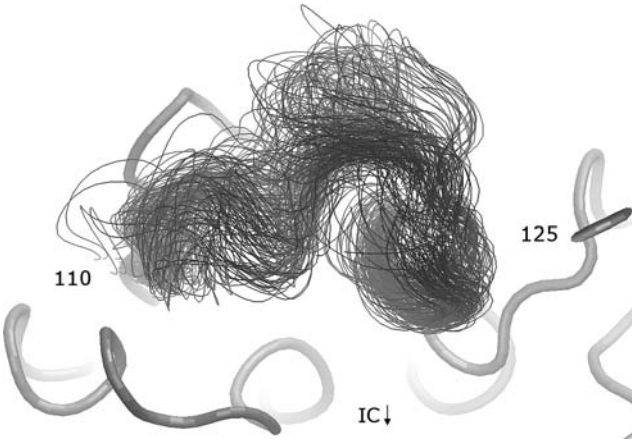

Fig. 6 Superimposition of 200 minimized frames from the high-temperature simulation. The frames were extracted from the 450 K trajectory and subsequently minimized. This picture was generated with a viewing axis perpendicular to the interdomain cleft. The location of residues 110 and 125 is indicated, as well as the direction of the cleft (''IC '')

Computational methods

Crystal lattices were built using DeepView [13]. The same program was used to analyze crystal contacts.

Molecular dynamics simulations and subsequent analyses of the trajectories were carried out using the AMBER suite of programs [14] and the AMBER94 force field [15]. Protein structures were obtained from the PDB, water molecules and other small compounds were removed using DeepView [13], and the simulation systems were prepared using the xleap utility. The Generalized Born model [16] was used to implicitly consider solvent effects on the simulation systems

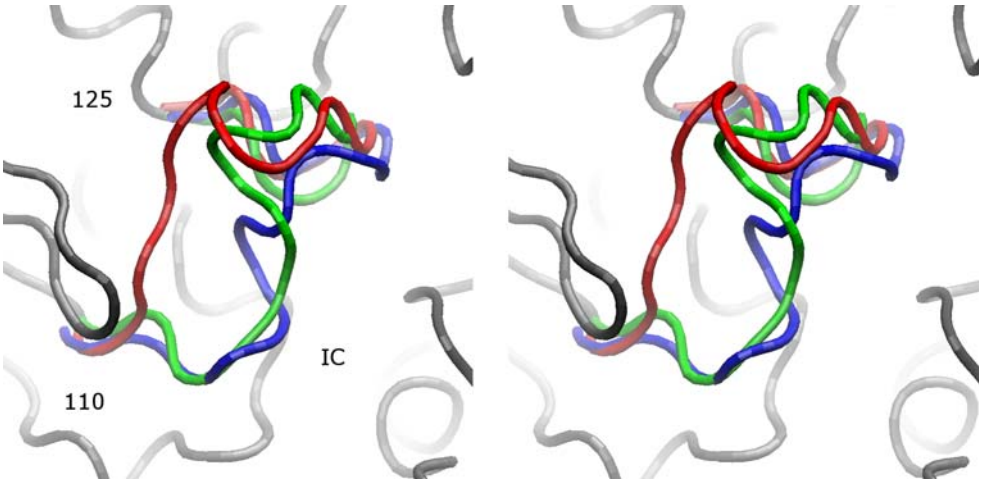

Fig. 7 (Stereo): Conformation of the Cys115 surface loop in representative members of the three conformational clusters obtained by high-temperature simulations. Shown in blue is a conformer of the largest cluster, the conformers from smaller clusters are red and green, respectively. The residue numbers denote the beginning and the end of the surface loop. The viewing axis is from the outside of the enzyme into the interdomain cleft (IC)

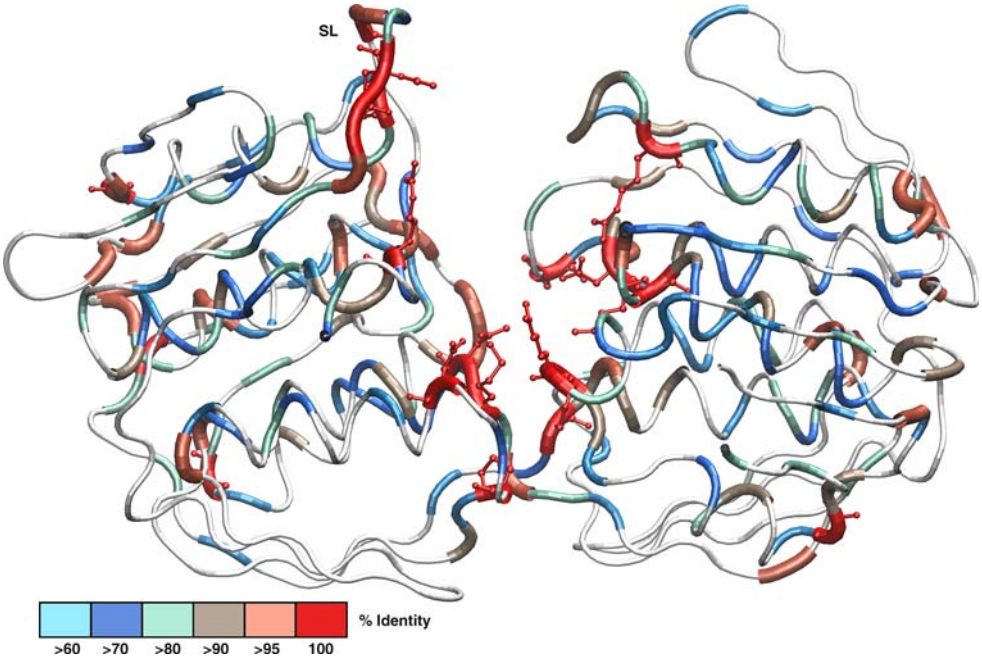

Fig. 8 Conserved and variable regions of MurA, shown here for the *E. cloacae* enzyme (pdb code 1NAW). The residues are colored according to their degree of conservation. Completely conserved residues are shown in red and with ball-and-stick side-chains. A VMD script that generates this representation for interactive viewing is available from the authors

while allowing a more efficient sampling of conformational space. Temperature regulation was done by coupling the simulation system to an external bath (time constant: 5 ps) [17]. The results were visualized using the VMD program [18]. The cluster analysis of the high-temperature MD results was performed using the g_cluster program of the Gromacs Package [19]. Further details of the MD simulation protocols are provides as Supporting Information.

Genome sequences of MurA were obtained from the ''UniProt Knowledgebase'' at http://www.expasy.ch/. The multiple alignment was computed using ClustalW [20]. Visualization and interactive analysis of sequence data were performed with the BioEdit program [21].

Acknowledgements The work has been performed under the Project HPC-EUROPA (RII3-CT-2003-506079), with the support of the European Community—Research Infrastructure Action under the FP6 ''Structuring the European Research Area'' Programme. This project was supported by the Deutsche Forschungsgemeinschaft (KL 1356) and the Fonds der Chemischen Industrie. We thank Professor R. Hartmann for his continued support of our work.

References

1. Brown ED, Vivas EI, Walsh CT, Kolter R (1995) J Bacteriol 177:4194
2. Skarzynski T, Mistry A, Wonacott A, Hutchinson SE, Kelly VA, Duncan K (1996) Structure 4:1465
3. Schonbrunn E, Sack S, Eschenburg S, Perrakis A, Krekel F, Amrhein N, Mandelkow E (1996) Structure 4:1065

4. Kahan F, Kahan J, Cassidy P, Kropp H (1974) Ann NY Acad Sci 235:364
5. Schonbrunn E, Eschenburg S, Luger K, Kabsch W, Amrhein N (2000) Proc Natl Acad Sci USA 97:6345
6. Eschenburg S, Schonbrunn E (2000) Proteins 40:290
7. Skarzynski T, Kim DH, Lees WJ, Walsh CT, Duncan K (1998) Biochemistry 37:2572
8. Eschenburg S, Kabsch W, Healy ML, Schonbrunn E (2003) J Biol Chem 278:49215
9. Eschenburg S, Priestman M, Schonbrunn E (2005) J Biol Chem 280:3757
10. Schonbrunn E, Eschenburg S, Krekel F, Luger K, Amrhein N (2000) Biochemistry 39:2164
11. Eschenburg S, Priestman MA, Abdul-Latif FA, Delachaume C, Fassy F, Schonbrunn E (2005) J Biol Chem 280:14070
12. Thomas AM, Ginj C, Jelesarov I, Amrhein N, Macheroux P (2004) Eur J Biochem 271:2682
13. Guex N, Peitsch MC (1997) Electrophoresis 18:2714
14. Weiner PK, Kollman PA (1981) J Comput Chem 2:287
15. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA (1995) J Am Chem Soc 117:5179
16. Tsui V, Case DA (2001) Biopolymers 56:275
17. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR (1984) J Chem Phys 81:3684
18. Humphrey W, Dalke A, Schulten K (1996) J Mol Graphics 14:33
19. Lindahl E, Hess B, van der Spoel D (2001) J Mol Mod 7:306
20. Thompson JD, Higgins DG, Gibson TJ (1994) Nucleic Acids Res 22:4673
21. Hall TA (1999) Nucl Acids Symp Ser 41:95