COMPUTATIONAL
ANDSTRUCTURAL
BIOTECHNOLOGY
J O U R N A L

Energy flow and intersubunit signalling in GSAM: A non-equilibrium molecular dynamics study

C. Harder-Viddal [a], R.M. Roshko [b], J. Stetefeld [c,d,e,f,*]

[a] Department of Chemistry and Physics, Canadian Mennonite University, 500 Shaftesbury Blvd, Winnipeg, Manitoba, Canada
[b] Department of Physics and Astronomy, University of Manitoba, 30A Sifton Rd, Winnipeg, Manitoba, Canada
[c] Department of Chemistry, University of Manitoba, 144 Dysart Rd, Winnipeg, Manitoba, Canada
[d] Center for Oil and Gas Research and Development (COGRAD), Canada
[e] Department of Biochemistry and Medical Genetics, University of Manitoba, Canada
[f] Department of Human Anatomy and Cell Science, University of Manitoba, Canada

A R T I C L E   I N F O

A B S T R A C T

Non-equilibrium molecular dynamics simulations of vibrational energy flow induced by the imposition of a thermal gradient have been performed on the μ2-dimeric enzyme glutamate-1-semialdehyde amino-mutase (GSAM), the key enzyme in the biosynthesis of chlorophyll, in order to identify energy transport pathways and to elucidate their role as potential allosteric communication networks for coordinating functional dynamics, specifically the negative cooperativity observed in the motion of the two active site gating loops. Fully atomistic MD simulations of thermal diffusion were executed with a GROMACS simulation package on a fully solvated GSAM enzyme by heating various active site target ligands (initially, catalytic intermediates and cofactors) to $300K$ while holding the remainder of the protein and the solvent bath at $10K$ and monitoring the temperature $T(t)$ of all the enzyme residues as a function of time over a $1ns$ observation window. Energy is observed to be deposited in a relatively small number of discrete chains of residues most of which contribute to specific structural or biochemical functionality. Thermal linkages between all thermally active chains were established by isolating a specific pair of chains and performing a thermal diffusion simulation on the pair, one held at $300K$ and the other at $10K$, with the rest of the protein frozen in its initial atomic configuration and thus thermally unresponsive. Proceeding in this way, it was possible to map out multiple pathways of vibrational energy flow leading from one of the active sites through a network of contiguous residues, many of which were evolutionarily conserved and linked by hydrogen bonds, into the other active site and ultimately to the other gating loop.

1. Introduction

The flow of energy through biomolecular systems and its relationship to system structure and biological function continues to be the subject of intense theoretical, experimental and computational investigation [1]. Much of this investigative effort has focused on energy conversion processes and energy transport in proteins, motivated by the multitude of different tasks which they execute [2]. Studies employing a variety of methods including ultrafast spectroscopies [3–16], nuclear magnetic resonance [17,18], multiple sequence analysis [19–21] and molecular dynam-

ics simulations [22–25] have highlighted the highly anisotropic nature of the energy flow and have identified pathways for the flow of vibrational energy that span many residues. Some of these transport channels appear to play a functional role in regulating allosteric communication between distant sites [17–20,26–29] and driving large amplitude, biochemically significant conformational changes which may involve the correlated motion of hundreds to thousands of atoms [30–32] or in funneling excess heat away from a reaction centre (usually a ligand binding or photoexcitation event) thereby preserving structural integrity and avoiding potential denaturation [33–40]. Other channels appear to be dead-ends with no apparent functional significance [2].

Classical fully-atomistic molecular dynamics simulations [22,23] are among the most widely used methodologies for exploring protein dynamics, including signaling pathways in proteins and

* Corresponding author at: Department of Chemistry, University of Manitoba, 144 Dysart Rd, Winnipeg, Manitoba, Canada.
  E-mail address: jorg.stetefeld@umanitoba.ca (J. Stetefeld).

nucleic acids [24,25]. It is possible to reconstruct energy flow channels through the tertiary structure of proteins by searching for correlations between residues in equilibrium MD simulations [41–43]. However, energy flux is inherently a non-equilibrium phenomenon, and more direct approaches are available through a variety of non-equilibrium MD techniques such as anisotropic thermal diffusion (ATD) [28,44,45], in which energy flow through coupled networks is induced by the local heating or cooling of selected atoms or residues, and pump–probe methods [29,46–48] which attempt to mimic ultraviolet excitation and infrared vibration pumping experiments [48–50]. However, in many cases, the analysis is complicated by extremely unfavourable signal-to-noise ratios which require the application of special strategies, like cooling parts of the system to unphysically low temperatures [28,44].

In the current investigation, nonequilibrium molecular dynamics simulations of vibrational energy flow induced by the imposition of a thermal gradient have been performed on the µ2-dimeric enzyme glutamate-1-semialdehyde aminomutase (GSAM) [51–55], the key enzyme in the biosynthesis of chlorophyll, in order to define energy transport pathways in this system and to elucidate their role as potential allosteric communication networks for coordinating functional dynamics. GSAM is a member of the $\alpha-$ family of vitamin B_6- dependent enzymes and is composed of two contiguous monomers, each consisting of 433 amino acid residues, a single active site which hosts the transformation of glutamate-1-semialdehyde (GSA) into 5-aminolevulinate (ALA), and an associated gating loop which undergoes dramatic conformational changes during catalysis and oscillates between an open configuration which allows substrate entry and product release and a closed configuration which obstructs access to the active site. In each subunit, the substrate GSA enters the active site when the gate is open, bonds with the pyridoxamine 5′-phosphate (PMP) form of the cofactor initiating the closure of the gate, evolves through a series of catalytic intermediates during which the cofactor is transformed into its pyridoxal 5′-phosphate (PLP) form, and is ultimately released as ALA when the gate reopens. However, the catalytic conversion sequences in the two subunits are out of phase and the two subunits exhibit negative cooperativity meaning that the oscillations of the two gates are synchronized in such a way that when the gate is open in one subunit it is simultaneously closed in the other. Such precise coordination clearly requires a highly efficient feedback loop for the rapid exchange of information between the two subunits. The present investigation seeks to establish potential candidates for this communication network by using the non-equilibrium MD methodology of anisotropic thermal diffusion (ATD) to generate a spatial map of thermal relaxation pathways through the protein residues following the deposition of excess thermal energy from a heat bath at selected target sites. The simulations show that the energy re-emerges preferentially in a relatively small number of thermally active chains of sequential residues, some of which lie on key architectural elements of the enzyme which are either integral to active site functionality and dynamics or which form part of the monomer–monomer interfacial (crossover) region.

2. Simulation methods

The starting point for the MD simulations was a high resolution crystal structure [55] obtained from X-ray diffraction measurements performed on a particular PMP/PLP form of the enzyme (PDB code 2HP2) which was reconstituted from the apoenzyme so as to trap specific catalytic intermediates in the two active sites, one containing the PMP-substrate complex bound in its ketamine-4 (KE4) form and with the gating loop in its open conformation, and the other containing an internal aldimine between PLP and Lys273 with DAVA as the intermediate and with the gating loop in its closed conformation. Fig. 1 shows a PyMOL image of GSAM

which highlights the principal structural features of the enzyme. These include the two monomeric subunits (A yellow and B white) each of which contributes an interface helix consisting of residues Ser122 to Gly139 (magenta), a gating loop consisting of residues Cys149 to Thr183 (cyan), and an active site containing substrate binding residues Ser28 to Gly39 and Glu406, cofactors PMP and PLP, catalytic intermediates KE4 and DAVA (red sticks), Asp245 for pyridine nitrogen fixation and residues Tyr301 to Thr305 responsible for fixing the phosphate group of the cofactor of the opposite subunit.

The simulations were performed with a GROMACS [56] package (version 4.5.4) using a standard force field ffG43a2 for the protein and an SPC model for the solvent. The equations of motion were integrated using a leapfrog algorithm with a time step $\Delta t = 1fs$. A particle mesh Ewald method was used to treat the long-ranged electrostatic interactions and the non-bonded (electrostatic and vdW) interaction pairs were updated every tenth step using a cut-off radius of $r = 1nm$. Internal topologies for each of the three non-protein molecules KE4, PLP and DAVA were prepared separately using PRODRG2 [57] and then incorporated manually into the overall system topology along with all of the necessary covalent bond, angle and dihedral connectivities to the protein amino acid residue atoms. The entire enzyme-substrate complex was solvated in a cubic box containing 51,660 water molecules and 23Na$^+$ ions and its energy was minimized by the method of steepest descent. Each simulation proceeded in three stages, (1) an initial relaxation for a period of 50ps in which the C_α backbone atoms were restrained to fixed positions, the pressure was held constant at $P = 1atm$ using a Berendsen pressure coupling algorithm with coupling time constant $\tau_P = 1.0ps$ and the temperatures of the solvent, protein, cofactors and catalytic intermediates KE4, PLP and DAVA were all held constant at $T = 10K$ using a Berendsen velocity-rescaling thermostat with coupling time constant $\tau_T = 0.1ps$, followed by (2) an unrestrained equilibration run at constant pressure $P = 1atm$ and constant temperature $T = 10K$ for a period of 10ps in which the position restraint on the C_α backbone atoms was lifted, and finally (3) a non-equilibrium production run for a period of 1ns at constant pressure $P = 1atm$ in which the catalytic intermediate or cofactor in one of the active sites (in this case, either KE4 or DAVA or PLP) was coupled to a heat bath at $T = 300K$ while the rest of the system continued to be held at a constant temperature $T = 10K$. (Unlike other investigations where the simulations were performed with the protein or nucleic acid complex decoupled from the solvent bath [28,44,45], thermal cou-
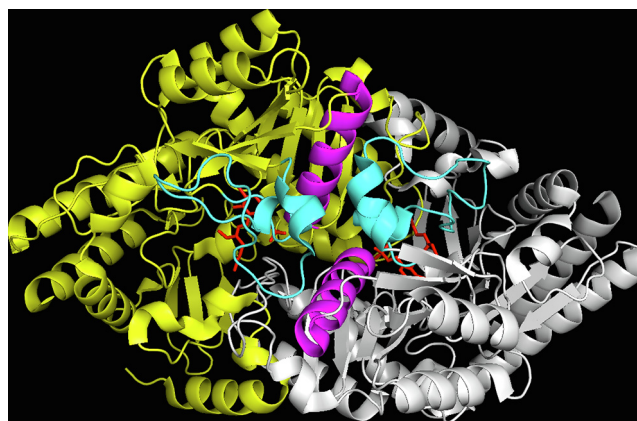
Fig. 1. A PyMOL image of GSAM which highlights the principal architectural features of the enzyme including the two monomeric subunits (A yellow and B white), the interface helices (magenta) and the gating loops (cyan). The catalytic intermediates KE4 and DAVA and the cofactor PLP are shown as red sticks. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

pling to the bath was maintained throughout the current simulations. While this procedure almost certainly compromised the efficiency of energy transport along functional communication pathways by diverting energy directly into the solvent, and suppressed the thermal responsiveness of the residues along the network, it also better reflects the real physical situation where the complex is immersed in an aqueous environment and where functionality must compete with non-functional losses.) An equilibrium background simulation was also performed in which all system components were held at $T = 10K$ throughout all three stages. Both simulation sequences were repeated 100 times in order to reduce fluctuations. The temperatures T_α of all of the individual amino acid residues and heterogeneous molecules were extracted from the simulation data at discrete intervals of $0.1ps$ and averaged over the 100 production runs by evaluating the defining relationship $(1/2)N_{df}^\alpha kT_\alpha = (1/2)\sum_{i=1}^{N_\alpha} m_{\alpha i} v_{\alpha i}^2$ where N_{df}^α is the number of degrees of freedom of molecule α, N_α is the number of atoms in molecule α, and $m_{\alpha i}$ and $v_{\alpha i}$ are the mass and velocity, respectively, of the ith atom in molecule α. Once the pattern of energy deposition was established and all of the thermally active chains of residues were identified from the completely unrestrained ATD simulations, a further sequence of restrained ATD simulations were performed in which the motion of the entire protein was frozen with the exception of two selected thermally active chains, one held at $300K$ and the other at $10K$. In this way, the energy transfer between pairs of architectural elements may be studied in isolation from the rest of the protein, thus yielding information about the possibility or impossibility of direct communication between any two elements and about the specific residues responsible for the communication.

The strategy underlying the decision to equilibrate the protein at a relatively low temperature $T = 10K$, well below normal range of physiological conditions, was designed to enhance the visibility of the thermal relaxation effects by reducing the noise due to thermal fluctuations [28,44]. Low temperature protocols like this have been criticized on the grounds that there may be insufficient thermal energy to allow an adequate sampling of conformation space structures. While a detailed comparison of several different ATD heating protocols performed on an RNA-ligand complex [44] does indeed reveal some differences in the response of specific nucleotides under a reversal of the thermal roles of the nucleic acid R and the ligand L from $R200K/L10K$ to $R10K/L200K$, many of the larger-scale systematic trends in the magnitude and time constant of the energy transfer are preserved, as are the principal conclusions regarding which elements act as communication "cornerstones". The effect of the simulation time step Δt on temperature changes derived from nonequilibrium simulations has been investigated in the literature [58] where it was found that a time step of $\Delta t = 1fs$ (used in the current study) could result in artificially high temperature changes and that a simulation step size $\Delta t = 0.7fs$ was optimal. In order to evaluate the consequences of this effect for the current investigation, additional simulations were generated with the smaller step size and the temperature changes of the residues were recorded. The thermal responsiveness ΔT (defined below) of all residues was found to decrease by a relatively small amount ($2K$ to $5K$) relative to the $1fs$ results, with no discernible changes in the pattern of energy deposition.

3. Anisotropic thermal diffusion in GSAM

### 3.1. Unrestrained ATD simulations: KE4 heated to 300 K

The principal features of the time dependence of the relaxation response obtained from the ATD simulations with the catalytic intermediate KE4 acting as the heat source are illustrated in

Fig. 2 in the form of plots of temperature T as a function of time on a logarithmic scale over four decades of observation time $0.1ps < t < 1000ps$ for the heat source KE4 (Fig. 2(a)) and for three representative residues, Thr124A (Fig. 2(b)) located on the interface helix in subunit A, His151A (Fig. 2(c)) located on the gating loop of subunit A and Gln302B (Fig. 2(d)) located in the active site of subunit A. Each figure includes a normative background response with KE4 held at the same temperature 10 K as the rest of the system. As a measure of the "thermal responsiveness" of a residue, we use the difference between the final temperature of the residue in the ATD simulation obtained by averaging over the last $100ps$ of the observation window and the background temperature averaged over the entire observation window, $\Delta T = < T_{ATD}(900ps < t < 1000ps) > - < T_{background}(0.1ps < t < 1000ps) >$. The thermal responsiveness ΔT is shown in Fig. 2 for all three residues.

Fig. 2(a) shows that the heat source KE4 stabilizes at its final equilibrium temperature approximately $1ps$ following initial contact with the $300K$ heat bath (with a thermostat coupling time constant $\tau_T = 0.1ps$). For the vast majority of residues, the transfer of energy from the heat source occurs very gradually over a much longer time scale and the time-averaged temperature rises slowly and monotonically from its initial equilibrium value (nominally $10K$) prior to heating. However, for a number of the most responsive residues (such as residue His151A on the gating loop shown in Fig. 2(c)), the energy transfer tracks the initial changes in the heat source and exhibits an initial rapid surge within the first few picoseconds with an amplitude which is well in excess of the thermal noise, before settling into a gradual monotonic upward relaxation. This surge is presumably a manifestation of a particularly advantageous side-chain orientation and coupling to the KE4 heat source. (Artificial peaks such as those which can sometimes be observed as a consequence of the release of position constraints have been eliminated from the production runs by performing all constraint releases in the $10K$ equilibration runs, prior to coupling KE4 to the $300K$ heat bath.)

In order to establish whether the process of thermal diffusion produces observable conformational distortions of the residues, the rmsd of every residue was calculated with reference to the final equilibrated structure, attained immediately prior to the injection of thermal energy into the active site, and plotted as a function of observation time over the interval $0.1ps < t < 1000ps$. Only three of the residues exhibited a clear response in the rmsd which was significantly in excess of the background rmsd obtained by holding the entire protein at 10 K, Asp155A on the gating loop of subunit A, and Tyr301B and Thr305B both of which are residues in subunit B which contribute to the fixation of the cofactor in the active site of subunit A. Fig. 3 shows the time dependence of the rmsd of these three residues with KE4 held at $300K$ (black curves) and, for comparison, the background rmsd in the absence of heating (green curves). In all three cases, the rmsd is initially coincident with the background but subsequently exhibits a relatively abrupt stepwise transition to a new time independent plateau. The transition occurs at t_{step} $10ps$ for the gating loop residue Asp155A (Fig. 3(a)) which has a centre of mass separation of d $1.48nm$ from KE4, and occurs earlier at t_{step} $1ps$ for the phosphate fixation residues Tyr301B (Fig. 3(b)) and Thr305B (Fig. 3(c)) both of which are closer to KE4 at separations of d $1.10nm$ and d $0.55nm$, respectively. While the effects described above are noteworthy inasmuch as they involve residues which are located on key functional groups of the enzyme, there are many more residues in these and other functionally and architecturally important groups in both subunits with stronger thermal relaxation responses but with much weaker or negligible changes in rmsd, so we conclude that conformational distortions alone are not reliable indicators of energy flow.
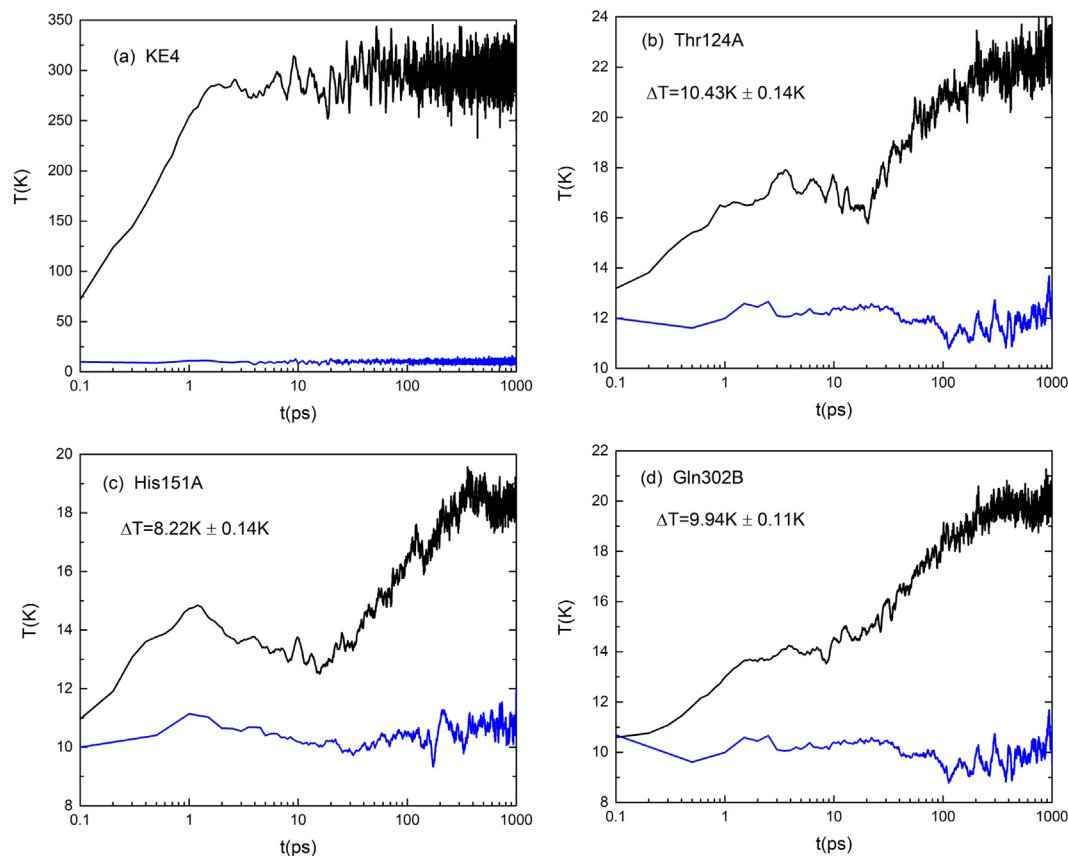
Fig. 2. The time dependence of the thermal relaxation response obtained from the ATD simulations in the form of plots of temperature T as a function of time on a logarithmic scale over four decades of observation time $0.1ps < t < 1000ps$ for (a) the heat source KE4 and three representative residues, (b) Thr124A located on the interface helix in subunit A, (c) His151A located on the gating loop of subunit A and (d) Gln302B located in the active site of subunit A. The data in blue shows the background response at 10 K. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 4 shows the thermal responsiveness profile for both subunits A and B of the GSAM enzyme in the form of a plot of ΔT as a function of residue index from $n = 1$ to $n = 866$. The cofactor PLP in subunit B has been assigned the index $n = 273$ since it is covalently attached to residue Lys273 in subunit B. Fig. 4 shows that the energy is deposited preferentially in eleven discrete chains of sequential, covalently coupled residues some of which lie on principal architectural/functional elements of the enzyme. Five of the chains (1, 6, 7, 8 and 11) contribute elements which are integral to the functionality of the active site of subunit A. Of the fourteen residues believed to be involved in substrate binding and enzyme catalysis within the active site pocket of subunit A (Ser28A to Gly39A, Lys273A and Glu406A), **chain 1** composed of residues Gly25A to Ala33A contributes Ser28A, Ser29A, Pro30A, Val31A, Arg32A, and Ala33A, **chain 7** composed of residues Leu268A to Gly286A contributes Lys273A for Schiff base linkage formation, and **chain 8** composed of residues Ala401A to Thr410A contributes Glu406A. **Chain 11** composed of residues Val294A to Thr319B contributes all 5 of the residues involved in the fixation of the phosphate group of the cofactor in subunit A, Tyr301B, Gln302B, Ala303B, Gly304B and Thr305B, while **chain 6** composed of residues Leu241A to Arg252A contributes Asp245A for pyridine nitrogen fixation. **Chain 3** is composed of residues Val117A to Ile133A and overlaps with a portion of the interface helix of subunit A (Ser122A to Gly139A). **Chain 4** is composed of residues Lys145A to Met156A and overlaps with a portion of the gating loop of subunit A (Cys149A to Thr181A). **Chain 10** is composed of residues Met116B to Ile133B and overlaps with a portion of the interface helix of subunit B (Ser122B to Gly139B). The remaining three

bands, **chain 2** (residues Ile61A to Ala70A), **chain 5** (residues Ile210A to Ser218A), and **chain 9** (residues Thr91B to Cys97B) do not appear to have a clear architectural or functional significance.

As a further aid to interpretation, it is convenient to group the residues into two broad categories according to the magnitude of their response, as either weakly to moderately responsive when $3K \leq \Delta T < 10K$ or as highly responsive when $\Delta T \geq 10K$. The latter category includes chain 3 residues Asn121A, Ser122A, Gly123A, Thr124A, Glu125A and Ala126A the last five of which are located at the N-terminal end of interface helix A, chain 4 residues Tyr150A, His151A, Gly152A, His153A and Ala154A all on gating loop A, chain 5 residue Asn217A, chain 7 residues Gly272A, Lys273A, Pro280A and Val 281A which includes active site (Schiff base linkage) residue Lys273A, chain 10 residues Asn121B and Glu125B both at the N-terminal end of interface helix B, and chain 11 residues Tyr301B, Gln302B, Ala303B, Gly304B, Thr305B, Leu306B, Ser307B, Gly308B, Asn309B and Pro310B, the first five of which belong to the active site of subunit A, with Gly304B and Thr305B being the two most responsive residues in the entire enzyme with $\Delta T > 20K$.

The data in Fig. 4 are replotted in Fig. 5 as a function of the distance d between the centre of mass of the residue and the centre of mass of the catalytic intermediate KE4. A preliminary examination of this figure shows that the energy transfer is highly anisotropic in the sense that ΔT is observed to fluctuate significantly by factors of up to 3 or 4 between residues for which the distance d from KE4 differs by $0.02nm$ or less. The anisotropy is most evident at the shortest distances d $0.5nm$ and diminishes steadily with increasing d, before eventually becoming indistinguishable from the thermal
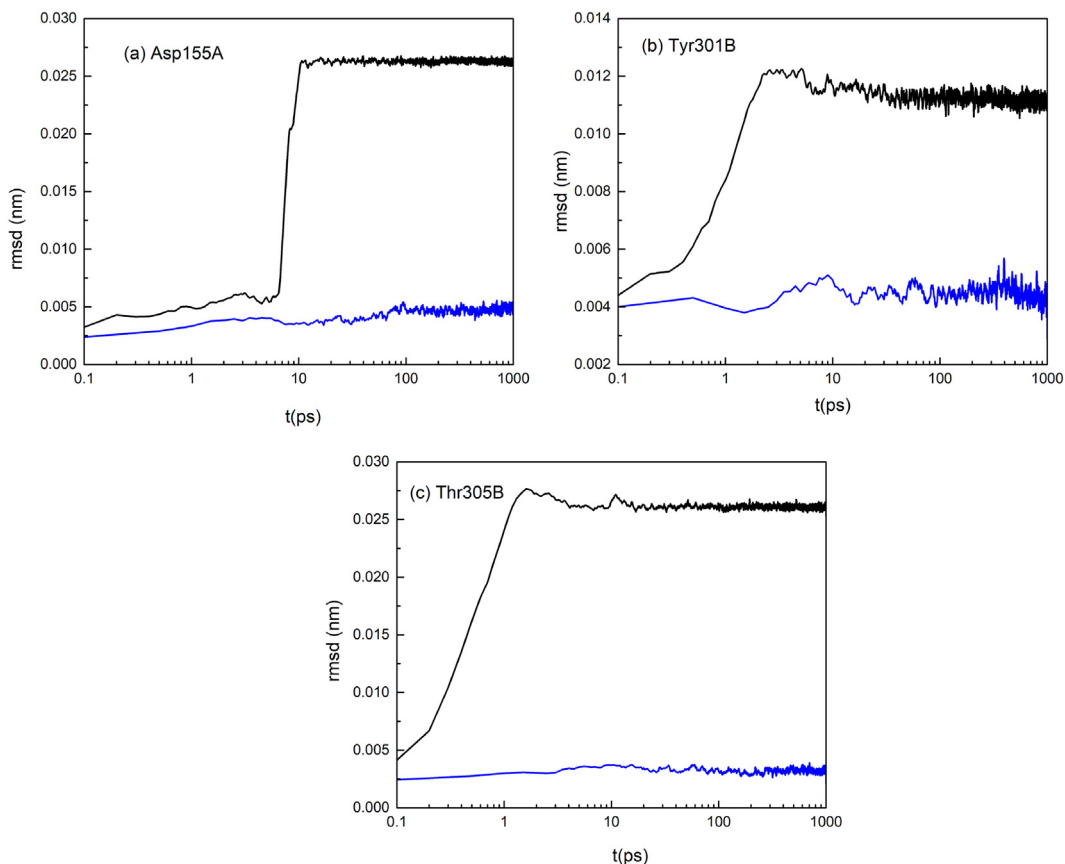
Fig. 3. The root mean square deviation (rmsd) calculated with reference to the final equilibrated structure attained immediately prior to the injection of thermal energy into the catalytic intermediate KE4 in active site A, and plotted as a function of observation time over the interval $0.1ps < t < 1000ps$ for three residues (a) Asp155A on the gating loop of subunit A and (b) Tyr301B and (c) Thr305B both of which are residues in subunit B which contribute to the fixation of the cofactor in the active site of subunit A. The green curves show the background rmsd at 10 K. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

noise above $d\ 2nm$. However, a closer inspection of Fig. 5 reveals that one chain of residues contributes to the anisotropy in a particularly unique way: chain 11 which consists entirely of residues

from subunit B (Val294B to Thr319B) not only contains some of the most responsive residues in the protein but these same residues also dominate the response on virtually every length scale where a competition exists with residues from other bands. In
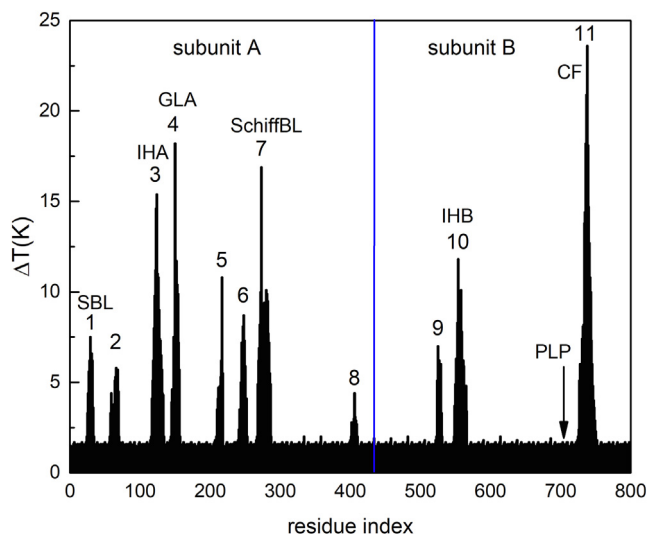
Fig. 4. The thermal responsiveness profile for all residues in both subunits A and B of the GSAM enzyme following the heating of the catalytic intermediate KE4 in subunit A, in the form of a plot of $\Delta T =< T_{ATD}(900ps < t < 1000ps) > - < T_{background}(0.1ps < t < 1000ps) >$ as a function of residue index from n = 1 to n = 866. The abbreviations are SBL = substrate binding loop, IH = interface helix, GL = gating loop, CF = cofactor fixation, SchiffBL = group that contains Lys 273 for Schiff base linkage formation.
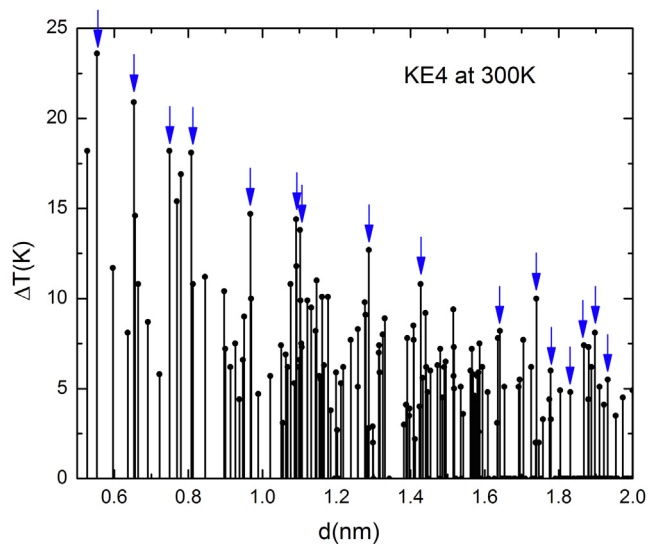
Fig. 5. The thermal responsiveness data in Fig. 4 replotted as a function of the distance d between the centre of mass of the residue and the centre of mass of the catalytic intermediate KE4.

Fig. 6. A PyMOL image of the pattern of thermal energy deposition on the six architecturally and functionally most significant chains of residues identified in Fig. 4 colour coded as follows: interface helices A and B (red), gating loop A (green), cofactor fixation group A (cyan), Schiff base linkage group A (blue), substrate binding loop A (lime), catalytic intermediate KE4 (yellow). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

order to emphasize this unique role, the residues belonging to chain 11 have been marked with vertical blue arrows in Fig. 5.

Fig. 6 shows A PyMOL image of the pattern of thermal energy deposition on the six architecturally and functionally most significant chains of residues identified in Fig. 4 and their relationship to the catalytic intermediate KE4 in subunit A, colour coded as follows: interface helices A and B (red), gating loop A (green), cofactor fixation group (cyan), Schiff base linkage group (blue), substrate binding loop A (lime), catalytic intermediate KE4 (yellow). Based on their location in the enzyme relative to the interfacial region, some the remaining chains appear to be either energetic "dead-ends" or thermal dissipation channels which are of questionable relevance to inter-subunit signaling and allosteric communication and are consequently omitted from the figure.

3.2. Unrestrained ATD simulations: PLP heated to 300 K

The ATD simulations were repeated with the cofactor PLP in subunit B coupled directly to the $300K$ heat bath. The general characteristics of the relaxation response are similar to those observed in the KE4-ATD analysis (see Fig. 2) and, once again, no consistent correlation was observed between the conformational distortion (rmsd) of a residue and energy flow as measured by its thermal responsiveness to heating DAVA. The thermal responsiveness profile shown in Fig. 7 is virtually the mirror image of the profile observed in the KE4-ATD analysis (see Fig. 4), which is consistent with the invariance of the molecular subunit structure of GSAM under a two-fold rotation. Thus, heating PLP excites the same eight bands of residues in subunit B that were excited in subunit A by heating KE4, and two of the three bands excited in subunit B by heating KE4 are also excited in subunit A by heating PLP. The single exception is the moderately strong band 9 in the KE4 analysis (Fig. 4), which was identified as an allosteric "dead-end" and which exhibits no thermal activity in response to heating PLP. The only other distinguishing feature of the PLP-ATD simulations is the very
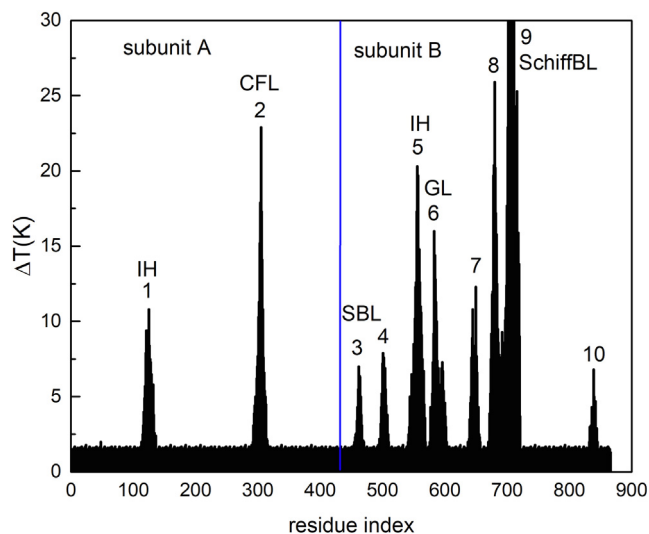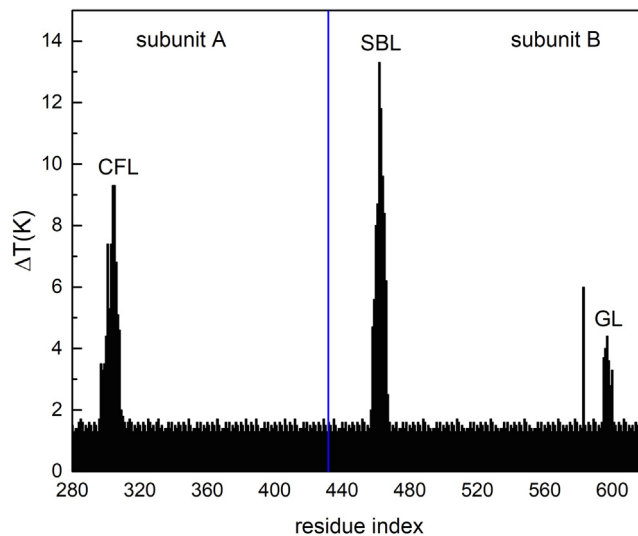

Fig. 7. The thermal responsiveness profile for all residues in both subunits A and B of the GSAM enzyme following the heating of the cofactor PLP in subunit B, in the form of a plot of ΔT as a function of residue index from n = 1 to n = 866. The abbreviations are SBL = substrate binding loop, IH = interface helix, GL = gating loop, CF = cofactor fixation, SchiffBL = group that contains Lys273 for Schiff base linkage formation.

high level of thermal activity observed throughout the entire chain of residues from 269B to 278B (the equivalent of band 7 in the KE4-ATD analysis) due to the covalent coupling of Lys273B directly to the PLP cofactor in subunit B. The thermal responsiveness of every residue in the chain exceeds any of those observed in the entire KE4-ATD simulation ($\Delta T > 30K$) and peaks at ΔT 170K in Lys273B.

3.3. Unrestrained ATD simulations: DAVA heated to 300 K

The ATD simulations were repeated once again with the catalytic intermediate DAVA in subunit B coupled directly to the 300 K heat bath and Fig. 8 shows the thermal responsiveness profile obtained from the simulations. According to this figure, energy transport from DAVA is limited to three functional groups of residues, the substrate binding loop (Gly25B to Ala33B), the cofactor

Fig. 8. The thermal responsiveness profile for all residues in both subunits A and B of the GSAM enzyme following the heating of the catalytic intermediate DAVA in subunit B, in the form of a plot of ΔT as a function of residue index from n = 1 to n = 866. The abbreviations are SBL = substrate binding loop, GL = gating loop, CF = cofactor fixation.

fixation loop (residues Tyr301A to Leu306A from subunit A), and the gating loop (Tyr150B and Gly162B to Val165B). A distinct feature of the DAVA-ATD simulations, which is absent from the both the KE4 and PLP simulations, is the potential communication link that exists between DAVA and the short helical section of the gating loop from Gly162B to Val165B which obstructs access to the active site in subunit B in the closed conformation. In the open conformation of the gate, this helical segment is transformed into a short twisted antiparallel $\beta-$ sheet.

3.4. Restrained ATD simulations

The unrestrained simulations in Section 3.3 show the pattern of energy deposition in GSAM following the local heating of a cofactor or catalytic intermediate. However, as a prerequisite to understanding allosteric communication, it is necessary to establish the precise pathways by which energy propagates through the protein by searching for linkages between the thermally active chains and distinguishing chain residues which are primary inter-chain receptors of energy from those which are secondary receptors in the sense that the energy they receive flows down the chain from the primary acceptors. A cross-correlation analysis of the relaxation response can provide information about the temporal sequence of events by which energy deposited at one site arrives at other sites in the protein [28,44]. However, the current investigation adopts a more heuristic approach in which the protein is systematically decomposed into pairs of elements each of which may be a single cofactor or a catalytic intermediate or an entire chain of residues of architectural significance, like an interface helix or a gating loop, or a subgroup thereof. Each member of the pair is free to move and to interact with the other member, while the position coordinates of the atoms in all of the remaining ele-

Table 1

Direct thermal linkages between functionally important groups of residues as determined by a restrained anisotropic thermal diffusion analysis.

ELEMENT 1 (300 K)	SUB UNIT	RESIDUES HEATED	ELEMENT 2 (10 K)	SUB UNIT	RESIDUES DIRECTLY COUPLED TO ELEMENT 1	$\Delta T(K)$
CATALYTIC INTERMEDIATE	A	KE4	INTERFACE HELIX	A	Ser122A	8
					Gly123A	10
					Thr124A	10
CATALYTIC INTERMEDIATE	A	KE4	GATE	A	Tyr150A	43
					His151A	37
					Gly152A	38
CATALYTIC INTERMEDIATE	A	KE4	COFACTOR FIXATION	A	Gly304B	41
					Thr305B	51
					Leu306B	42
CATALYTIC INTERMEDIATE	A	KE4	SUBSTRATEBINDING LOOP	A	Ser29A	41
CATALYTIC INTERMEDIATE	A	KE4	SCHIFF BASE LINKAGE	A	Lys273A	22
INTERFACE HELIX	A	Ser122A	INTERFACE HELIX	B	Ser122B	21
		Gly123A			Glu125B	24
		Thr124A				
		Glu125A				
INTERFACE HELIX	B	Asn121B	COFACTOR	B	PLP	30
		Ser122B			Pro280B	11
		Gly123B			Val281B	14
		Thr124B			Gly282B	13
		Glu125B			Ala283B	9
GATING LOOP	A	His153A	INTERFACE HELIX	B	Ser122B	10
		Ala154A			Gly123B	11
		Asp155A			Thr124B	12
		Met156A			Glu125B	14
					Ala126B	13
					Cys127B	14
					Met128B	19*
					Ala129B	15
					Val130B	16
					Leu131B	17
					Arg132B	26*
					Ile133B	15
					Met134B	15
					Arg135B	11
					Ala136B	13
					Tyr137B	6
					Thr138B	10
					Gly139B	10
COFACTOR FIXATION	A	Ala303B	INTERFACE HELIX	B	Val120B	14
		Gly304B			Asn121B	18*
		Thr305B				
		Leu306B				
		Ser307B				
CATALYTIC INTERMEDIATE	B	DAVA	GATING LOOP	B	Ser163B	30
					Gly164B	65
CATALYTIC INTERMEDIATE	B	DAVA	COFACTOR FIXATION	B	Ala303A	32
					Gly304A	41
					Thr305A	51
					Leu306A	42
					Ser307A	32
CATALYTIC INTERMEDIATE	B	DAVA	SUBSTRATEBINDING LOOP	B	Ser28B	28
					Ser29B	35
					Pro30B	32
					Val31B	27
					Arg32B	21

Table 2
Hydrogen bond connectivities between functionally important groups of residues.

ELEMENT 1	RESIDUE	ELEMENT 2	RESIDUE	PERSISTENCE LEVEL (%)
CATALYTIC INTERMEDIATE A	KE4	INTERFACE HELIX A	Gly123A	50
	KE4		Thr124A	5
CATALYTIC INTERMEDIATE A	KE4	COFACTOR FIXATION A	Thr305B	100
CATALYTIC INTERMEDIATE A	KE4	SCHIFF BASE LINKAGE A	Lys273A	50
CATALYTIC INTERMEDIATE A	KE4	CRYSTAL WATERS	W7367	100
GATING LOOP A	Tyr150A	CRYSTAL WATERS	W7367	100
GATING LOOP A	His153A	INTERFACE	Arg132B	100
	Asp155A	HELIX B	Arg132B	100
INTERFACE HELIX A	Ser122A	INTERFACE HELIX B	Glu125B	100
	Glu125A		Ser122B	100
INTERFACE HELIX B	Thr124B	COFACTOR B	PLP	100
COFACTOR FIXATION A	Ser307B	INTERFACE HELIX B	Asn121B	20
CATALYTIC INTERMEDIATE B	DAVA	COFACTOR FIXATION B	Gly304A	95
CATALYTIC INTERMEDIATE B	DAVA	SUBSTRATEBINDING LOOP B	Ser29B	100
			Arg32B	98

ments of the protein are frozen in a rigid spatial configuration. If one member of the chosen pair of elements is heated to 300K while the other is held at 10K, then it is possible to study in detail the coupling between any given pair of elements in isolation from the rest of the protein and thus to identify potential primary pathways for energy conduction between functionally significant regions within a given subunit and also between subunits.

Table 1 summarizes all the pairs of elements that were investigated. The first 3 columns of each row contain the element that acted as the heat source (element 1), the subunit A or B to which it belongs and the specific residues that were heated to 300K, and the last 4 columns contain the element that was held at 10K (element 2), the subunit to which it belongs, the specific residues that received energy directly from the first element and the thermal responsiveness $\Delta T(K)$ of these residues after 1ns.

The procedure can be illustrated with reference to the first row of the table. In the initial configuration, the entire system was frozen with the exception of the catalytic intermediate KE4 in subunit A and the entire interface helix in subunit A consisting of the 18 residues Ser122A, Gly123A, Thr124A, Glu125A, Ala126A, Cys127A, Met128A, Ala129A, Val130A, Leu131A, Arg132A, Ile133A, Met134A, Arg135A, Ala136A, Tyr137A, Thr138A, Gly139A. Heating KE4 to 300 K yielded a measurable response ($\Delta T > 3K$) in the eight residues Ser122A, Gly123A, Thr124A, Glu125A, Ala126A, Cys127A, Met128A of interface helix A. The response of the remaining ten residues was indistinguishable from the background response at 10K within the noise. In order to isolate which of these eight residues received energy directly from KE4 and which received energy indirectly by propagation down the chain from the primary receptors, selected subsets of the eight active responders were frozen until the smallest frozen subset that completely extinguished the response of all of the remaining residues was identified. In this way, it was established that the three residues Ser122A, Gly123A, Thr124A were each coupled directly to KE4 while the remaining five residues Glu125A, Ala126A, Cys127A, Met128A were not. Subsequent heating of interface helix A to 300K to establish thermal linkages to other architectural elements, for example to interface helix B at 10K, was then restricted to heating the group of three primary receptors Ser122A, Gly123A and Thr124A on helix A, since the thermal responsiveness of the remaining five secondary receptors diminished rapidly with distance down the chain from the primary group, making them unlikely candidates for energy transmission to other chains.

3.5. Hydrogen bond connectivities

Since networks of persistent H-bonds are believed to play an important role in the physical framework underlying signal propa-

gation [30], an analysis of H-bond couplings between the principal architectural elements of GSAM was performed and the results of this analysis are listed in Table 2. The first and third columns of each row contain the two architectural elements which share at least one H-bond, the second and fourth columns contain the specific residues which are linked by the H-bonds, and the last column shows the persistence level of the bond, defined as the fraction of the complete simulation time that at least one H-bond exists.

A comparison of Tables 1 and 2 shows that H-bonds are present in many, but not all, of the thermal linkages identified in the restrained ATD analysis in Section 3.4 above. Thus, the catalytic intermediate KE4 (between PMP and ALA) in subunit A is H-bonded directly to Thr305B (cofactor fixation), Gly123A and Thr124A (interface helix A) and Lys273A, and indirectly to Tyr150A (gating loop A) via the crystal water W7367, but is not H-bonded to Ser29A (substrate binding). The catalytic intermediate DAVA in subunit B is H-bonded directly to Gly304A (cofactor fixation) and to Ser29B and Arg32B (substrate binding) but not to Ser163B and Gly164B (helical section of gating loop B). The interface helix and gating loop of subunit A, both of which exhibit a direct thermal linkage to the interface helix element of subunit B, are also hydrogen bonded directly to the interface helix B: interface helix A (Ser122A → Glu125B, Glu125A → Ser122B) and gating loop A (His153A → Arg132B, Asp155A → Arg132B). The catalytic intermediate KE4 in subunit A, which is not directly linked thermally to interface helix B but is linked indirectly to helix B via pathways through cofactor fixation group A and gate A, as shown by the restrained ATD simulations, is also linked indirectly to interface helix B by two independent chains of H-bonds, one via cofactor fixation A KE4 → Thr305B, Ser307B → Asn121B) and another via gate A (KE4 → W7367 → Tyr150A, His153A → Arg132B, Asp155A → Arg132B). Fig. 9 shows a PyMOL image of the hydrogen bond connectivities between all of the chains of residues which exhibit a response to the local injection of thermal energy into the actives sites of GSAM (see Figs. 4, 7 and 8).

3.6. Evolutionarily conserved residues

The observation in non-equilibrium MD simulations of discrete subsets of residues which are excited by a local injection of thermal energy into the catalytic intermediates within the active sites of GSAM and their interpretation as a potential system of allosteric signaling pathways linking functionally important sites within and between subunits is strengthened by other analyses which establish energetic connectivity and functional coupling between sites by means of evolutionary data [19–21]. A multiple sequence alignment (MSA) performed on a family of 86 GSAM enzymes col-
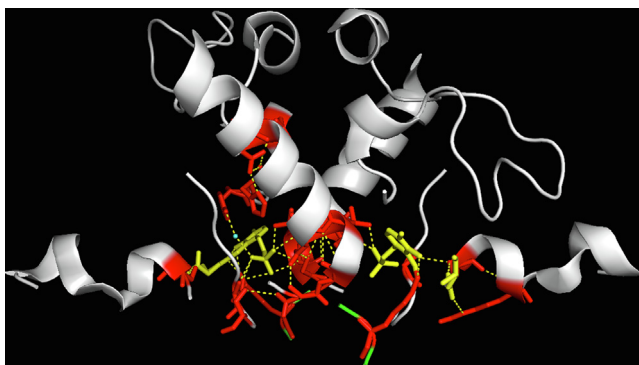
Fig. 9. A PyMOL image showing the hydrogen bond connectivities between all of the chains of residues which exhibit a response to the local injection of thermal energy into the actives sites of GSAM (see Figs. 4, 7 and 8).

lected from the nonredundant database using PSI-BLAST [59] and initially aligned using ClustalW [60] shows that 71 of the 433 amino acid residues in each subunit are (completely) evolutionarily conserved and thus contribute to either enzymatic folding or function. The MD simulations presented here show that 60 of the 141 conserved residues, 41 in subunit A and 19 in subunit B are excited by the injection of heat into the catalytic intermediate KE4 in subunit A, while 68 of the conserved residues, 18 in subunit A and 50 in subunit B, are excited by the injection of heat into the cofactor PLP in subunit B. In the former case where KE4 is the heat source, the list of thermally active conserved residues includes the five substrate binding residues Ser29A, Pro30A, Val31A, Arg32A and Glu406A, the same six residues Ser122A/B, Gly123A/B, Thr124A/B, Glu125A/B, Ala126A/B and Arg132A/B on both interface helices A and B, the four residues Tyr150A, His151A, Gly152A and His153A on gating loop A, the five cofactor fixation residues Tyr301B, Gln302B, Ala303B, Gly304B and Thr305B in subunit B as well as residue Asp245A for pyridine nitrogen fixation and the residue Lys273A for Schiff base linkage formation. A similar list can be constructed for the case where PLP acts as the heat source and includes all of the residues in the previous list but in the opposite subunit with several additions, specifically, the same five substrate binding residues, the same six residues on both interface helices A and B plus Arg135 on both helices, the same four residues on the gating loop plus Leu158B, Gly162B, Ser163B and Thr167B also on the gating loop, the same five cofactor fixation residues as well as Asp245B and Lys273B.

3.7. Discussion and conclusions

The identification of discrete subsets of residues which respond to a local injection of thermal energy into the catalytic intermediates or cofactors within the active sites of GSAM, using non-equilibrium MD simulations of thermal diffusion, shows that the architecture of GSAM potentially supports a number of specific pathways for the flow of vibrational energy within and between the two subunits of the enzyme. Assuming that the thermal excitation with a heat bath acts as a physical proxy for chemical activity, these pathways may form the basis of a communication network for coordinating and triggering allosteric conformational changes in the two subunits by shuttling information about chemical modifications of the catalytic intermediates within the two active sites back and forth between the two subunits.

It is important to place the current investigation in perspective with respect to some of the previous investigations that have employed a variety of theoretical/computational methods to identify and catalogue energy transport channels in biomolecules.

Guided by statistical analyses of evolutionary data that proposed the existence of an allosteric molecular switch consisting of a system of juxtaposed evolutionarily conserved residues, Ota and Agard [28] introduced the method of anisotropic thermal diffusion to identify an intramolecular communication pathway in the PDZ domain of the signaling protein PSD-95 by tracking changes in the rmsd down a chain of residues following heating of a single key residue responsible for ligand specificity. The entire protein was held at ultra-cold temperatures, and the ATD simulations were performed on the PSD-95-ligand complex in the gas phase by imposing weak harmonic restraints on the backbone atoms. The signal propagation was attributed primarily to van der Waals interactions between side-chain residues along the path. Nguyen et al. [44] performed similar non-equilibrium MD simulations of thermally induced energy flow on the aptamer domain of a riboswitch-guanine complex. The time evolution of the kinetic energy of each nucleotide was used to track the energy flow, which was observed to spread anisotropically through a hydrogen-bonding network connecting the guanine ligand to the surrounding nucleotides. Stacking interactions played only a minor role in the signal propagation. The entire complex was equilibrated at $200K$ and the thermal relaxation data were collected by coupling the guanine ligand to a heat bath at $10K$ and simultaneously disconnecting all other nucleotides from the $200K$ bath. Martinez et al. [45] heated each individual residue in the ligand binding domain of the $\beta-$ subtype thyroid hormone receptor in separate simulations and measured the resulting thermal energy deposited in the remaining, artificially cooled, protein environment. A comparison was performed with the results obtained using the same protocol but with each residue mutated to a glycine or an alanine. The difference in thermal energies between the two systems allowed the major heat diffusers to be identified, and these same residues have also been identified experimentally as playing a key functional role in transactivation. Comparative analyses of both dynamically responsive protein regions and specific hydrogen-bonding patterns in different domains of two structural homologs of the heat shock protein Hsp70 family of chaperones – one allosteric (Hsp70-DnaK) and the other non-allosteric (Hsp110-Sse1) -- have been performed by Chiappori et al. [30] as a way of determining a molecular mechanism for the propagation of an allosteric signal. In particular, the hydrogen-bonded network that bridges the ligand-binding site and the substrate binding domain in the allosteric protein appear to be modulated by the presence of the ligand, in that hydrogen bonds persist for a greater percentage of the total simulation time with the ligand bound. No ligand-modulation was observed for the non-allosteric protein. MD simulations by Ishikura et al. [43] were used to calculate inter-residue energy conductivities in order to analyze energy flow patterns that play a key role in the long-range intramolecular cross-talk between distant regions of the photosensory receptor photoactive yellow protein (PYP). By evaluating the energy density at each site after a long simulation, the energy conductivity between the chromophore and the surrounding amino acid residues was obtained. The residues with the largest energy conductivities were indicative of a primary energy transfer pathway between the two locations. A subset of the residues with large conductivities formed a hydrogen-bond network with the chromophore, enabling the transfer of energy from the chromophore.

The simulation protocols employed in the current investigation differ in certain respects from those described above. As mentioned earlier, all of the simulations presented here were performed over a $1ns$ observation time window with the protein subject to the natural conformational restraints imposed by a fully solvated environment (as opposed to the artificial constraints required by simulations in vacuo) while simultaneously and continuously maintaining thermal contact with two thermal baths: the hot
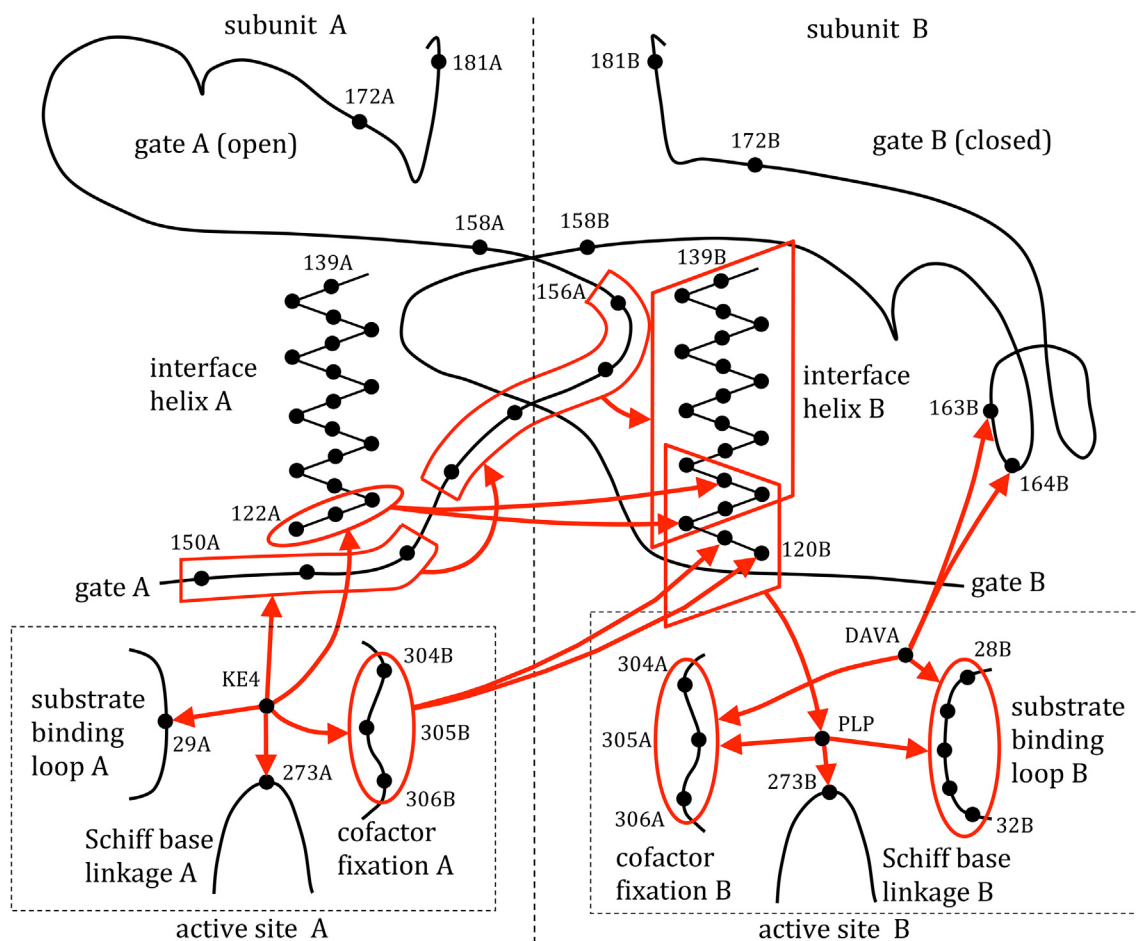
Fig. 10. A schematic depiction of the energy flow pathways between the principal architectural and functional elements in the GSAM enzyme based on the connectivities listed in Table 1 as well as the unrestrained simulations in Fig. 7 (PLP heated) and Fig. 8 (DAVA heated).

ligand with a 300*K* bath and the remainder of the protein plus the solvent with a 10*K* bath. The latter restriction was necessary in order to avoid artificial thermal drifts which were observed in background simulations (all system components held at 10*K*) whenever thermal contact with the cold bath was broken for times exceeding about 50*ps*. These drifts may not have been observable in the previous investigations which were typically performed over shorter time scales $t < 100ps$. The current procedure also incorporates natural dissipation pathways that undoubtedly influence the efficiency of energy transport in real systems where it is likely that some energy is diverted away from functional tasks directly into the solvent.

Another feature which distinguishes the current simulations from previous studies is the method used to reconstruct potential transport channels from the pattern of energy deposition established by the heating of the cofactor or the catalytic intermediate. In order to determine which of the discrete chains of residues shown in Figs. 4, 7 and 8 receive energy directly from the catalytic intermediate/cofactor and which receive energy indirectly through other thermally active chains, the entire protein was rigidly frozen in its initial set of atomic coordinates with the exception of the catalytic intermediate/cofactor and one thermally active chain. An ATD simulation was then performed by heating the catalytic intermediate/cofactor to 300*K* and observing the thermal response of the selected chain which was simultaneously held at 10*K*. In this way, it was found that a subset of the thermally active chains received no energy directly from the catalytic intermediate/cofactor. If a given chain was found to be thermally active, then a further

sequence of ATD simulations were repeated on that chain by systematically freezing individual residues or groups of residues in the chain until each chain was decomposed into a "primary" subset of residues that received energy directly from the catalytic intermediate/cofactor and a "secondary" subset that received energy indirectly only from the primary residues by propagation down the chain. Subsequent ATD simulations then focused on establishing inter-chain linkages by choosing a specific pair of chains and repeating the above process, with one member of the pair designated as the heat source and the other member as the heat sink, while the remainder of the protein was held frozen.

Coarse-grained approaches [61,62], which avoid the significant statistical noise and lengthy simulation times of the fully-atomistic techniques employed in the current investigation, have also been developed to identify energy transport networks in proteins. These approaches use a normal-mode harmonic approximation, which neglects long wavelength modes, to describe dynamics over long time scales from a master equation which predicts the time-dependent probability for a specific residue. The master equation is generated with constants for energy transfer between residue pairs, based on energy diffusion constants between them. A network of weighted local transport coefficients throughout the protein can then be constructed, which acts as a map of the energy diffusion pathways. A comparison of the fully-atomistic nonequilibrium diffusion approach and the coarse-grained normal mode approach with a master equation was recently performed on the 36-residue headpiece domain of villin [63]. The authors concluded that there was a high degree of similarity between the two

approaches, suggesting that a relatively large protein like GSAM would be a good candidate for the simpler coarse-grained method in future studies.

While the studies mentioned in the comparisons above are generally supportive of a connection between vibrational energy transport and allostery in proteins and view rapid vibrations as a necessary precursor to slower barrier activated dynamics and a useful predictor [28,64] of large conformational changes characteristic of allosteric transitions or structural changes in molecular motors [65–67], other studies question the validity of a relationship based on such widely disparate time scales and believe that further investigation of the extent and limitations of such connections is required [68–70].

Finally, it is particularly relevant to remark on several recent studies of vibrational energy flow and its relationship to signaling pathways in other allosteric proteins which share a dimeric structural similarity with GSAM. Equilibrium molecular dynamics simulations have been performed on the oxygen-sensor protein FixLH [71] in order to simulate ligand (O_2) binding in the dimeric sensor domain. The interaction between the sensor domain and the kinase domain is known to play a major role in signal transductions in two-component systems that detect physical and chemical stimuli in their environment [72]. The strength of inter-residue interactions was evaluated from molecular dynamics simulations by calculating the change in magnitude of the energy-transport coefficients relative to the deoxy model. The largest change in these coefficients occurred for residues on the interface between the two sensor subunits as well as at the C-terminal end of the two subunits both of which are linked to the kinase domain through a coiled-coil segment, suggesting that these two regions are key to signal transduction. A similar conclusion is reached in the current study on GSAM, performed using non-equilibrium molecular dynamics. The $\alpha-$ helices on the dimeric interface of GSAM were identified as preferential responders to energy flow following ligand heating, suggesting that the interface region is crucial to inter-subunit signaling. The gate of each subunit, which closes upon substrate entry and opens upon product release, also responded strongly to energy flow. However, in contrast to FixLH, in which the coiled-coil linker connected to the highly responsive C-terminal transmits signals to the kinase domain, for GSAM, the gates themselves change conformation.

The results of a systematic and comprehensive analysis of vibrational energy flow involving all pairs of thermally active chains conducted in the current study is summarized in Fig. 10 which shows a schematic depiction of the energy flow pathways between the principal architectural and functional elements in the GSAM enzyme based on the freeze group connectivities listed in Table 1 as well as the unrestrained simulations in Fig. 7 (PLP heated) and Fig. 8 (DAVA heated). The arrows in this figure originate on the hot element and terminate on the cold element. Inspection of this figure reveals an asymmetry in the energy flow patterns obtained by heating the catalytic intermediate KE4 in subunit A and heating the equivalent catalytic intermediate DAVA in subunit B. KE4 communicates directly with the substrate binding loop, the cofactor fixation residues, the Schiff base linkage residue, the interface helix and the rigid portion of the gating loop which lies outside the hinge residues Leu158A and Ser172A, while DAVA communicates only with the substrate binding loop, the cofactor fixation residues and the flexible lid portion of the gating loop between the hinge residues, specifically with residues Ser163B and Gly164B on the short helical section of the gate which obstructs access to the active site. Further inspection shows that all signals that originate somewhere within the active site of subunit A (from catalytic intermediate KE4, substrate binding loop, Schiff base linkage or cofactor fixation loop) and potentially terminate somewhere in subunit B must pass through both interface

helix B and the cofactor PLP in the active of subunit B. This can obviously occur through multiple pathways some of which are direct (cofactor fixation A [Gly304B, Thr305B, Leu306B] → interface helix B [Val120B, Asn121B] → PLP) and some of which are indirect (catalytic intermediate KE4 → gate A primary [Tyr150, His151, Gly152] → gate A secondary [His153, Ala154, Asp155, Met156] → interface helix B [Ser122B to Gly139B] → PLP, KE4 → interface helix A [Ser122A, Gly123A, Thr124A] → interface helix B [Ser122B, Glu125B] → PLP). Once the energy has reached PLP, the local connectivities in subunit B (see Fig. 10) allow the energy to flow to the cofactor fixation [Gly304A, Thr305A, Leu306A] and substrate binding residues [Ser28B, Ser29B, Pro30B, Val31B, Arg32B], from there to the catalytic intermediate DAVA, and hence ultimately to the helical section of gate B [Ser163B, Gly164B]. In this way, configurational changes associated with the enzymatic reaction steps in the active site of subunit A can eventually be communicated to the gating loop in subunit B, triggering a conformational transformation in the helical segment to a $\beta-$ sheet and the subsequent opening of the gating loop.

Future work will investigate the effect of mutations on the energy flow patterns, specifically, how the energy conduction pathways are altered by mutating the most significant responders among the evolutionary conserved residues. Such studies will reveal the extent to which the side-chains of the individual residues transmit energy, ultimately providing insight into the physical significance of the evolutionarily conserved residues. Several aspects of the mechanics of gate opening and closing will also be investigated. MD steering and umbrella sampling techniques will be applied to extract the potential of mean force, which provides a profile of the free energy barriers that the gates encounter as they open and close. X-ray diffraction data shows that the conformation of the gate differs in secondary structure over the short segment involving residues 159 to 171. In the open form of the gate, where the cofactor has the PMP form, the stabilizing element is a short, twisted, antiparallel $\beta-$ sheet secondary structure while, in the closed form of the gate, where the cofactor has the PLP form, a short subset of these residues from 164 to 168 exhibits a helical structure. MD simulations performed over a 7ns time window show that these open-gate-PMP and closed-gate-PLP conformations are indeed very stable. However, preliminary MD simulations show that, if the cofactors in the two subunits are exchanged, then the presence of PLP in the open-gate-subunit appears to trigger a conformational change within the first 100ps in the secondary structure involving residues 164 to 168 from a $\beta-$ sheet to a helix, indicating the presence of a communication pathway between the cofactor and the gate. By contrast, the presence of PMP in the closed-gate-subunit did not appear to alter the gate helix. Future simulations will explore these observations in more detail.

The role of hydrogen bond connectivity in facilitating thermal diffusion in GSAM will also be investigated in more detail. Recent studies [73,74] have shown that equilibrium fluctuations in the length of the hydrogen bonded contact have an impact of the rate of vibrational energy transfer which was found to vary inversely with the variance of the contact distance, so that a rigid contact with a small variance mediates vibrational motion essentially instantaneously while a floppy contact with a large variance corresponds to a slow transport pathway. In principle, the freeze group technique used in the current investigation to study interactions between pairs of functionally important groups of residues in isolation from the rest of the protein can be refined to create pairs of individual residues, one acting as a 300 K heat source and the other acting as a 10 K sink for energy flow. In this way it would, for example, be possible to compare the thermal relaxation of two neighbouring residues Gly123 and Thr124 on interface helix A, both of which receive energy directly from the catalytic intermedi-

ate KE4 in the active site of subunit A and both of which have contact hydrogen bonds with KE4 with variances of $0.04nm^2$ and $0.006nm^2$, respectively. However, when a T-jump excitation is employed, the ramping time can be significant and care must be taken to freeze all degrees of freedom of the target residues receiving the energy and to release these degrees of freedom by restoring the velocities of the frozen atoms to their original 10 K equilibrium values only after the source residue has reached its final temperature. Otherwise, thermal relaxation response functions will contain a significant initial transient component corresponding to vibrational energy delivered during the ramping process which obscures the arrival time and distorts the shape of the response. Even if such precautions are taken, it can still be very difficult to observe the initial arrival of the vibrational energy and the initial shape of the target response using a T-jump excitation. A related question concerns the quantifying of the role played by the small interfacial water cluster of the dimer by determining the rate of vibrational energy transport to it. In the current study, one water molecule in the cluster was found to have a persistent hydrogen-bond with KE4, allowing it to receive and transmit energy across the cluster and beyond. Previous studies of vibrational energy transport rates on homodimeric hemoglobin (HbI) [75,76] have shown that small water clusters between the dimer interfaces enhance energy transport, as they receive energy at a higher rate from locally hot regions due to the higher thermal conductivity of water relative to amino acids, thus playing an essential role as an energy conduit in the biomolecule.

The simulations presented here suggest a possible molecular scaffolding for intersubunit communication which is responsible for the negative cooperativity between the gating loops in the two allosteric monomers in GSAM. However, the reasons why the enzyme performs these collective, large amplitude, conformational changes involving the motion of hundreds of atoms remain elusive. The answer may lie in the formation of a ternary complex with glutamyl-tRNA reductase (GluTR) which channels the substrate between two subsequent enzymes in the tetrapyrrole biosynthesis, and experimental efforts to co-crystallize the GluTR-GSAM complex are currently underway.

Acknowledgements

References

[1] Leitner DM, Straub JE. Proteins: energy, heat and signal flow. CRC Press; 2009.
[2] Leitner DM. Energy flow in proteins. Annu Rev Phys Chem 2008;59:233–59.
[3] Nagy AM, Raicu V, Miller RJD. Nonlinear optical studies of hemeprotein dynamics: implications for proteins as hybrid states of matter. Biochim Biophys Acta 2005;1749:148–72.
[4] Kholodenko Y, Volk M, Goodig E, Hochstrasser RM. Energy dissipation and relaxation processes in deoxymyoglobin after photoexcitation in the Soret region. Chem Phys 2000;259:71–87.
[5] Lian T, Locke B, Kholodenko Y, Hochstrasser RM. Energy flow from solute to solvent probed by femtosecond IR spectroscopy: malachite green and heme protein solutions. J Phys Chem 1994;98:11648–56.
[6] Mizutani Y. Time-resolved resonance Raman spectroscopy and application to studies on ultrafast protein dynamics. Bull Chem Soc Japan 2017;90:1344–71.
[7] Lim M, Jackson TA, Anfinrud PA. Femtosecond near-IR absorbance study of photoexcited myoglobin: dynamics of electronic and thermal relaxation. J Phys Chem 1996;100:12043–51.
[8] Simpson MC, Peterson ES, Shannon CF, Eads DD, Friedman JM, et al. Transient Raman observations of heme electronic and vibrational photodynamics in deoxyhemoglobin. J Am Chem Soc 1997;119:5110–7.
[9] Ye X, Demidov A, Champion PM. Measurements of the photodissociation quantum yields of MbNO and MbO2 and the vibrational relaxation of the six-coordinate heme species. J Am Chem Soc 2002;124:5914–24.
[10] Ye X, Demidov A, Rosca F, Wang W, Kumar A, et al. Investigations of heme absorption line shapes, vibrational relaxation, and resonance Raman scattering on ultrafast time scales. J Phys Chem A 2003;107:8156–65.
[11] Hummer G, Schotte F, Anfinrud PA. Unveiling functional protein motions with picosecond X-ray crystallography and molecular dynamics simulations. PNAS 2004;101:15330–4.
[12] Li P, Champion PM. Investigations of the thermal response of laser-excited biomolecules. Biophys J 1994;66:430–6.
[13] Sagnella DE, Straub JE, Jackson TA, Lim M, Anfinrud PA. Vibrational population relaxation of carbon monoxide in the heme pocket of carbonmonoxy myoglobin: comparison of time-resolved mid-IR absorbance experiments and molecular dynamics simulations. PNAS 1999;96:14324–9.
[14] Botan V, Backus EHG, Pfister R, Moretto A, Crisma M, et al. Energy transport in peptide helices. PNAS 2007;104:12749–54.
[15] Kukura P, McCamant DW, Mathies RA. Femtosecond stimulated Raman spectroscopy. Annu Rev Phys Chem 2007;58:461–88.
[16] Mukherjee P, Kass I, Arkin IT, Zanni MT. Picosecond dynamics of a membrane protein revealed by 2D IR. PNAS 2006;103:3528–33.
[17] Clarkson MW, Lee AL. Long-range dynamic effects of point mutations propagate through side chains in the serine protease inhibitor eglin c. Biochemistry 2004;43:12448–58.
[18] Fuentes EJ, Gilmore SA, Mauldin RV, Lee AL. Evaluation of energetic and dynamics coupling networks in a PDZ domain protein. J Mol Biol 2006;364:337–51.
[19] Lockless SW, Ranganathan R. Evolutionarily conserved pathways of energetic connectivity in protein families. Science 1999;286:295–9.
[20] Schulman AI, Larson C, Mangelsdorf DJ, Ranganathan R. Structural determinants of allosteric ligand activation in RXR heterodimers. Cell 2004;116:417–29.
[21] Suel GM, Lockless SW, Wall MA, Ranganathan R. Evolutionarily conserved networks of residues mediate allosteric communication in proteins. Nat Struct Biol 2003;10:59–69.
[22] Van Gunsteren WF, Weiner PK, Wilkinson AJ. Computer simulation of biomolecular systems: theoretical and experimental applications. Springer; 1997.
[23] Karplus M, McCammon JA. Molecular dynamics simulations of biomolecules. Nat Struct Biol 2002;9:646–52.
[24] Feher VA, Durrant JD, Wart ATV, Amaro RE. Computational approaches to mapping allosteric pathways. Curr Opin Struct Biol 2014;25:98–103.
[25] Guo J, Zhou HX. Protein allostery and conformational dynamics. Chem Rev 2016;11:6503–15.
[26] Swain JF, Gierasch LM. The changing landscape of protein allostery. Curr Opin Struct Biol 2006;16:102–8.
[27] Formaneck MS, Ma L, Cui Q. Reconciling the "old" and "new" view of protein allostery: a molecular simulation study of chemotaxis Y protein (CheY). Proteins 2006;63:846–67.
[28] Ota N, Agard DA. Intramolecular signaling pathways revealed by modeling anisotropic thermal diffusion. J Mol Biol 2005;351:345–54.
[29] Sharp K, Skinner J. Pump-probe molecular dynamics as a tool for studying protein motion and long range coupling. Proteins 2006;65:347–61.
[30] Chiappori F, Merelli I, Colombo G, Milanesi L, Morra G. Molecular mechanism of allosteric communication in Hsp70 revealed by molecular dynamics simulations. PLoS Comput Biol 2012;8:e1002844.
[31] Nicolai A, Delarue P, Senet P. Decipher the mechanisms of protein conformational changes induced by nucleotide binding through free-energy landscape analysis: ATP binding to Hsp70. PLoS Comput Biol 2013;9:e1003379.
[32] Grant BJ, Gorfe AA, McCammon JA. Large conformational changes in proteins: signaling and other functions. Curr Opin Struct Biol 2010;20:142–7.
[33] Henry ER, Eaton WA, Hochstrasser RM. Molecular dynamics simulations of cooling in laser-excited heme proteins. PNAS 1986;83:8982–6.
[34] Koyama M, Neya S, Mizutani Y. Role of hemepropionates of myoglobin in vibrational energy relaxation. Chem Phys Lett 2006;430:404–8.
[35] Bu L, Straub JE. Vibrational energy relaxation of 'tailored' hemes in myoglobin following ligand photolysis supports energy funneling mechanism of heme 'cooling'. J Phys Chem B 2003;107:10634–9.
[36] Sagnella DE, Straub JE. Directed energy "funneling" mechanism for heme cooling following ligand photolysis or direct excitation in solvated carbonmonoxy myoglobin. J Phys Chem B 2001;105:7057–63.
[37] Sagnella DE, Straub JE, Thirumalai D. Timescales and pathways for kinetic energy relaxation in solvated proteins: application to carbonmonoxy myoglobin. J Chem Phys 2000;113:7702–11.
[38] Takayanagi M, Okumura H, Nagaoka M. Anisotropic structural relaxation and its correlation with the excess energy diffusion in the incipient process of photodissociated MbCO: high-resolution analysis via ensemble perturbation method. J Phys Chem B 2007;111:864–9.
[39] Bu L, Straub JE. Simulating vibrational energy flow in proteins: relaxation rate and mechanism for heme cooling in cytochrome c. J Phys Chem B 2003;107:12339–45.
[40] Lampa-Pastirk S, Beck WF. Intramolecular vibrational preparation of unfolding transition state of Zn^{II}-substituted cytochrome c. J Phys Chem B 2006;110:22971–4.
[41] Ghosh A, Vishveshwara S. A study of communication pathways in methionyl-tRAN synthetase by molecular dynamics simulations and structure network analysis. PNAS 2007;104:15711–6.
[42] Kong Y, Karplus M. Signaling pathways of PDZ2 domain: a molecular dynamics interaction correlation analysis. Proteins 2009;74:145–54.

[43] Ishikura T, Yamato T. Energy transfer pathways relevant for long-range intramolecular signaling of photosensory protein revealed by microscopic energy conductivity analysis. Chem Phys Lett 2006;432:533–7.

[44] Nguyen PH, Derreumaux P, Stock G. Energy flow and long-range correlations in guanine-binding riboswitch: a nonequilibrium molecular dynamics study. J Phys Chem 2009;113:9340–7.

[45] Martinez L, Figueira ACM, Webb P, Polikarpov I, Skaf MS. Mapping the intramolecular vibrational energy flow in proteins reveals functionally important residues. J Phys Chem Lett 2011;2:2073–8.

[46] Nguyen PH, Park S, Stock G. Nonequilibrium molecular dynamics simulation of the energy transport through a peptide helix. J Chem Phys 2010;132:025102.

[47] Kobus M, Nguyen PH, Stock G. Coherent vibrational energy transfer along a peptide helix. J Chem Phys 2011;134:124518.

[48] Nguyen PH, Stock G. Nonequilibrium molecular dynamics simulation of a photoswitchable peptide. Chem Phys 2006;323:36–44.

[49] Nguyen PH, Stock G. Chem Phys 2003;119:11350.

[50] Backus E, Nguyen PH, Botan V, Pfister R, Moretto A, Crisma M, et al. J Phys Chem B 2008;112:9091–9.

[51] Jordan PM, Shemin D. J Biol Chem 1973;248:1019–24.

[52] Pugh CE, Harwood JL, John RA. J Biol Chem 1992;267:1584–8.

[53] Smith MA, Kannangara CG, Grimm B. Biochemistry 1992;31:11249–54.

[54] Hennig M, Grimm B, Contestabile R, John RA, Jansonius JN. PNAS 1997;94:4866–71.

[55] Stetefeld J, Jenny M, Burkhard P. Intersubunit signaling in glutamate-1-semialdehyde-aminomutase. Proc Nat Acad Sci 2006;103:13688–93.

[56] Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ. GROMACS: fast, flexible, and free. J Comput Chem 2005;26:1701–18.

[57] Schüttelkopf AW, van Aalten DMF. PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. Acta Crystallogr D 2004;60:1355–63.

[58] Gulzar et al. Energy transport pathways in proteins: a non-equilibrium molecular dynamics simulation study. J Chem Theory Comput 2019;15:5750–7.

[59] Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, et al. Nucleic Acids Res 1997;25:3389–402.

[60] Thompson JD, Higgins DG, Gibson TJ. Nucleic Acids Res 1994;22:4673–80.

[61] Leitner DM, Pandey HD, Reid KM. Energy transport across interfaces in biomolecular systems. J Phys Chem B 2019;123:9507–24.

[62] Leitner DM and Yamato, T. "Mapping energy transport networks in proteins," Rev. Comp. Chem.; Parrill, A. L., Lipkowitz, K. B., Eds.; John Wiley & Sons, Inc. (2018); Vol. 31; pp. 63 - 114;

[63] Leitner DM, Buchenberg S, Brettel P, Stock G. Vibrational energy flow in the villin headpiece subdomain: master equation simulations. J Chem Phys 2015;142:075101.

[64] Smock RG, Gierasch LM. Sending signals dynamically. Science 2009;324:198–203.

[65] Lee Y, Choi S, Hyeon C. Mapping the intramolecular signal transduction of G-protein coupled receptors. Proteins Struct Funct Genet 2014;82:727–43.

[66] Thirumalai D, Hyeon C. Signalling networks and dynamics of allosteric transitions in bacterial chaperonin groel: implications for iterative annealing of misfolded proteins. Philos Trans R Soc B 2018;373:20170182.

[67] Hyeon C, Onuchic JN. Energy balance and dynamics of kinesin motors. In Proteins: energy, heat and signal flow; Leitner, D. M., Straub, J. E., Eds.; CRC Press, Taylor & Frances Group: Boca Raton, FL, 2010; pp 3−21.

[68] Stock G, Hamm PA. Nonequilibrium approach to allosteric communication. Philos Trans R Soc B 2018;373:20170187.

[69] Lu C, Knecht V, Stock G. Long-range conformational response of a PDZ domain to ligand binding and release: a molecular dynamics study. J Chem Theory Comput 2016;12:870–8.

[70] Buchenberg S, Sittel F, Stock G. Time-resolved observation of protein allosteric communication. PNAS 2017;114:E6804–11.

[71] Ota, Yamato. Energy exchange network model demonstrates protein allosteric transition: an application to an oxygen sensor protein. J Phys Chem B 2019;123:768–75.

[72] Stock AM, Robinson VL, Goudreau PM. Two-component signal transduction. Annu Rev Biochem 2000;69(1):183–215.

[73] Reid KM, Yamato T, Leitner DM. Scaling of rates of vibrational energy transfer in proteins with equilibrium dynamics and entropy. J Phys Chem B 2018;122:9331–9.

[74] Buchenberg S, Leitner DM, Stock G. Scaling rules for vibrational energy transport in proteins. J Phys Chem Lett 2016;7:25–30.

[75] Gnanasekaran R, Agbo JK, Leitner DM. Communication maps computed for homodimeric hemoglobin: computational study of water-mediated energy transport in proteins. J. Chem. Phys. 2011;135. 0065103.

[76] Reid KM, Yamato T, Leitner DM. Variation of energy transfer rates across protein-water contacts with equilibrium structural fluctuations of a homodimeric hemoglobin. J Phys Chem B 2020;124:1148–59.