Biochemistry

 Volume 35, Number 25 June 25, 1996

New Concepts in Biochemistry

A Mechanism for Reducing Entropic Cost of Induced Fit in Protein−RNA Recognition†

Lluís Ribas de Pouplana, Douglas S. Auld, Sunghoon Kim,‡ and Paul Schimmel*

Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Received February 1, 1996; Revised Manuscript Received May 1, 1996⊗

ABSTRACT: Induced fit has been postulated to be an important component of ligand interactions with proteins, including protein−DNA interactions. We imagined that the entropic cost of induced fit might be highly dependent on the local protein sequence context around critical contact residues. To investigate this question, we analyzed the basis for active or inactive phenotypes found in a library of combinatorial sequence variants of a surface-located helix−loop peptide which is essential for the anticodon-binding activity of a class I tRNA synthetase. Molecular dynamics simulations of the domain encompassing the helix−loop peptide of the active variants consistently demonstrated fixation of the local motion of five critical (for function) residues which are highly mobile in inactive variants. Additional experiments with other rationally chosen mutants extended the correlation between phenotype and motion of these vital residues. We propose that the need for fixation of local motion is an important constraint on sequences of surface peptides which form parts of RNA-binding sites. The fixation of motion of critical residues in the unbound protein can significantly reduce the entropic cost of complex formation by induced fit.

Protein−nucleic acid recognition is based on specific direct or indirect (water-mediated) contacts between bases and functional groups of amino acids (Vinson et al., 1989; Steitz, 1990; Harrison, 1991; Burd & Dreyfuss, 1994). Typically, the directed replacement of a contact residue results in a weakening or disruption of the protein−nucleic acid complex (Lim et al., 1991; Pabo & Sauer, 1992; Tan et al., 1993). Less understood is the role of noncontact residues and of sequence context, where loss of function cannot be ascribed to a loss of the potential for a specific contact or to denaturation, but rather to subtle effects of sequence context on the details of the three-dimensional arrangement and motion of atoms at the site where contacts normally would occur. This concept is related to that of coupling local folding to the binding of nucleic acids, as described by Spolar and Record (1994). This coupling of local folding with binding amounts to induced fit as first presented by Koshland (1958). We decided to investigate specifically this aspect of recognition, using a relatively well-studied system as an experimental model.

Crystallographic analyses of complexes of tRNAs with their cognate tRNA synthetases have provided a molecular picture of the detailed contacts needed for specific protein−tRNA recognition (Rould et al., 1989; Ruff et al., 1991; Biou et al., 1994). In general, major contacts are seen with the tRNA acceptor stem and, in two cases studied in some detail, with the anticodon (Rould et al., 1989; Ruff et al., 1991). Methionyl- and isoleucyl-tRNA synthetases are class I tRNA synthetases which share a common anticodon-binding motif centered around a 30 amino acid helix−loop peptide (Figure 1A,B; Brunie et al., 1990). The anticodon-binding helix−

† This work was supported by Grants GM15539 and GM23562 from the National Institutes of Health and by a grant from the National Foundation for Cancer Research.

* To whom correspondence should be addressed: Tel 617-253-6739; FAX 617-253-6636.

‡ Present address: Department of Biology, Sung Kyun Kwan University, Suwon, South Korea.

⊗ Abstract published in *Advance ACS Abstracts*, June 1, 1996.

S0006-2960(96)00256-5 CCC: $12.00

loop peptide of the three-dimensional crystal structure of *Escherichia coli* methionyl-tRNA synthetase is also a model for the analogous motif of the isoleucine enzyme. The significance of this motif for anticodon recognition was shown by experiments which used a single amino acid swap in a designed peptide to switch recognition between anticodons of $tRNA^{Met}$ and $tRNA^{Ile}$ (Auld & Schimmel, 1995). These anticodons, in turn, differ by just one nucleotide [CAU (Met) versus GAU (Ile)].

Although the cocrystal structure of neither enzyme with its bound tRNA has been solved, extensive mutagenesis of the 30 amino acid helix−loop peptide of methionyl-tRNA synthetase revealed the importance of a 14 amino acid subregion extending from Asn452 to Lys465 (Kim & Schimmel, 1992; Kim et al., 1993a−c, 1994). In one series of experiments, combinatorial variants of the 30 amino acid motif were generated, and active variants were selected for their capacity to complement a Δ*metG* null strain (Kim et al., 1993c). This strain has a complete ablation of the chromosomally encoded gene for methionyl-tRNA synthetase. The library was designed to give binary combinations at each position of the wild-type amino acid replaced by either alanine (for a hydrophobic wild-type residue) or serine (for a hydrophililc residue). The library had a complexity of about 2×10^7, and active variants were scored with a frequency of roughly 1%. Sequences of active and inactive variants showed that active variants had as many as 19 alanines and serines out of 30 residues. Further analysis and mutagenesis suggested that Asn452, Arg453, Pro460, Trp461, and Lys465 were immutable, regardless of sequence context (Kim et al., 1994). These five residues are positioned on the same side of the structure where they lie at the end of the helix and within the loop (Figure 1).

All other positions in the helix−loop peptide were substituted in one or more of the active variants. The occurrence of a particular substitution at a specific position of the sequence in active proteins clearly depended on sequence context. However, the explanation for which sequences would result in active or inactive proteins was obscure. For example, many inactive variants retained the five immutable residues and seemed stable as judged by their accumulation *in vivo* and by the ability to be purified and characterized *in vitro*. For this reason we decided to search for an alternative explanation for the inactivity of mutant proteins containing the five required amino acids. As an initial attempt, we tried to correlate the sequence-dependent changes in intramolecular contact energies (van der Waals and electrostatic interactions) with the activities of the variants. No clear correlation could be found between loss of side-chain interactions in the models and their active or inactive nature.

We then focused on the possibility that, while many inactive mutants were stable and seemingly had all residues needed for a favorable interaction energy with the RNA ligand, inactivation occurred as a consequence of the higher entropic cost of making a well-defined, relatively fixed protein−RNA interface. This higher cost could arise from the need to "immobilize" critical surface residues when they contacted the RNA. If the unbound residues in the inactive variants experienced a high degree of motion, then the entropic cost of making the complex (with residue motion now constrained) would be relatively high, when compared to a complex formed with the same residues in the context of a sequence which greatly constrained critical residue motion in the unbound state. To pursue this line of analysis, we turned to molecular dynamics simulations of residue motion (Karplus & Petsko, 1990; van Gunsteren & Mark, 1992). We placed particular emphasis on the effect of sequence variations on the motion of critical residues whose identities were maintained in the active and inactive variants.

Three-dimensional models were constructed from several of the MetRS[1] sequences in the combinatorial library and then studied by simulating dynamic motion for periods of 50−100 ps. The trajectories from each simulation were then analyzed for correlations between the dynamic behavior of a model and its active or inactive nature. Once several sequences from the combinatorial library had been studied, models carrying new mutations were built. These models had mutations not previously investigated and, as a consequence, provided specific predictions which could be tested by *in vivo* complementation experiments.

EXPERIMENTAL PROCEDURES

Combinatorial Mutagenesis Library and Complementation Assay. The construction and analysis of the combinatorial library of MetRS have been described previously (Kim et al., 1994). Site-directed mutagenesis to construct new mutants was according to Kim et al. (1993b). The tester strain Δ*metG*(MN9261/pRMS615) was used in complementation assays to select active variants of MetRS (Kim et al., 1993c). The production and the stability of the variants *in vivo* were determined by immunological methods, as described earlier (Shepard et al., 1992).

Model Building, Refinement, and Molecular Dynamics (MD) Simulations. Because of the large size of MetRS and the large number of models to be analyzed, it was necessary to reduce the complexity of the system. The presence of two distinct domains in the protein offered the possibility of dissecting the structure into two independent pieces. All energy minimization and molecular dynamics calculations were done using X-PLOR (version 3; Brünger et al., 1987) on a Silicon Graphics Indigo RS4400 workstation. The empirical energy function of X-PLOR and its standard parameter files were used to calculate the motions of residues in all of the models over total simulation times ranging from 95 to 145 ps (45 ps of equilibration time plus 50 or 100 ps of productive simulation). SHAKE constraints (Ryckaert et al., 1977) with a relative geometric tolerance of 0.0005 Å were used in all hydrogen−heavy atom bonds. The overall simulation process consisted of the following steps:

(a) Model Building and Refinement. For each sequence a three-dimensional model was created by placing the appropriate amino acid substitutions into the crystallographic model of the structure of MetRS (Brunie et al., 1990). Those coordinates constituted the starting structure for the energy refinement and dynamics simulations. All calculations from this point were carried out identically for both wild-type MetRS and the models.

The initial coordinates were minimized in the following manner. First, energy minimization using conjugate gradients (2000 cycles) was performed to remove large steric

[1] Abbreviations: MetRS, methionyl-tRNA synthetase; IleRS, isoleucyl-tRNA synthetase; Amp, ampicillin; Cmp, chloramphenicol; MD, molecular dynamics; SDS−PAGE, sodium dodecyl sulfate−polyacrylamide gel electrophoresis; IPTG, isopropyl β-D-thiogalactoside.

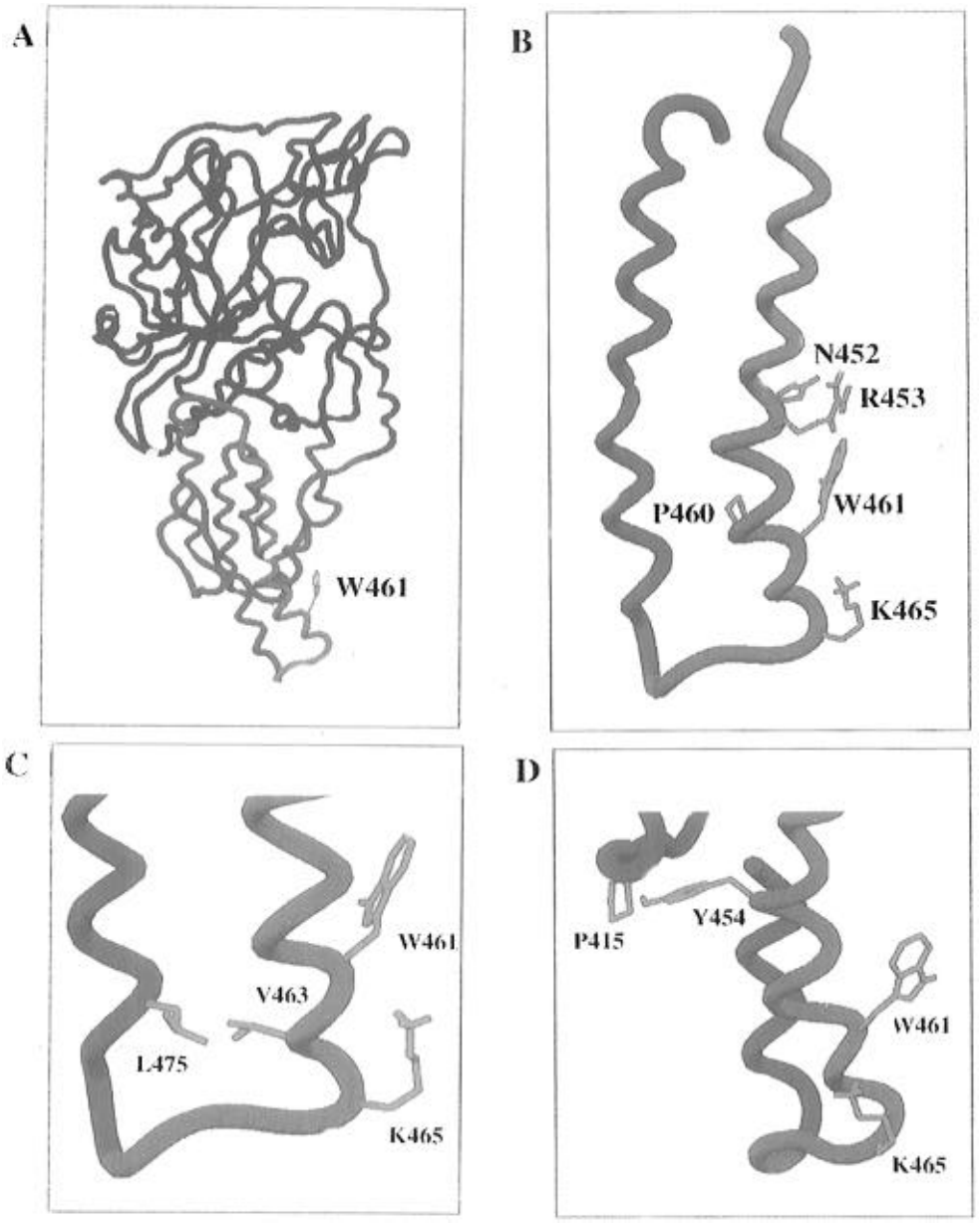

FIGURE 1: (A) Representation of the backbone of the three-dimensional structure of MetRS (Brunie et al., 1990). The active site domain is at the top of the molecule and is depicted in black. The anticodon-binding domain is located at the bottom and is shown in blue. The anticodon-binding peptide analyzed in this work is highlighted in red. The side chain of the essential W461 is shown in green. (B) Enlarged backbone view of the anticodon-binding peptide of MetRS, from residues 439 to 490. The region mutagenized in the combinatorial library (Kim et al., 1993c) is shown in red. The residues found to be essential in the context of the highly mutated sequence AV16 (Kim et al., 1994, and Table 1) are depicted and labeled. (C) View of the packing interaction between V463 and L473 at the core of the loop structure connecting the two helices of the studied peptide. The disruption of this interaction has a dramatic effect on mobility of the region and results in an inactive enzyme in the context of a mutagenized sequence. (D) View of the packing interaction between P415 and Y454. The disruption of this interaction by substituting P415 for glycine does not result in loss of activity nor in perturbation of the motions of critical residues.

clashes. At that point SHAKE constraints were introduced and side-chain charges of Asp, Glu, Lys, and Arg were set to zero in order to carry out the simulation protocol *in vacuo*. The temperature of the system was raised to 1000 K and then cooled to 0 K, by decreasing the temperature in 40 steps of 25 K a piece. At each temperature level, the dynamic motions of the model were calculated for 250 integration steps of 0.1 fs each. The final coordinates obtained at 0 K were finally refined again with conjugate gradient methods for 5000 cycles or until an energy gradient norm of less than 0.01 was reached. From this point SHAKE constraints for all hydrogen−heavy atom bonds were used.

In a further analysis, calculations were limited to the anticodon-binding domain alone, by calculating motion for residues 370−510. Interactions of this domain with the active site domain were simulated by including residues 364−369 and 511−515 in the calculations, but keeping fixed their α-carbon coordinates. These residues lie at the interface between the two domains of the protein and were chosen to take account of the interfacial domain−domain interactions. Previously, we had established that this approximation resulted in a large decrease in computational time without causing appreciable changes in the residue motions in the anticodon-binding domain (see Results).

(b) System Equilibration. The first step of the dynamics simulation consisted of a thermalization step by which the temperature of the system was raised from 0 to 300 K in 60 steps of 5 K each. At each temperature step the model was

allowed to equilibrate for 0.2 ps, using 200 integration steps of 1 fs. This equilibration was followed by 30 cycles of simulation at 300 K (0.4 ps per cycle, 1 fs integration time), with reassignment of the system's temperature to 300 K at the beginning of each cycle. The last equilibration process before productive dynamics was a 30 000 step simulation (1 fs integration time, 30 ps total time), with rescaling of the velocities of the system every 400 steps to achieve a temperature of 300 K.

(c) Productive Dynamics. The dynamic trajectories for all of the models were recorded as the final step of the calculations. This step consisted of 50 000 or 100 000 time steps of 1 fs, at a temperature of 300 K with initial velocities assigned to a Maxwellian distribution using the X-PLOR program IASVEL. The coordinates of the models were recorded at the end of each picosecond of simulation. No temperature scaling or velocity reassignment method was used during this phase of the simulation. The fluctuation of total energy and temperature of the system versus the simulation time for each MD run was monitored. Typically, fluctuations of these values were less than ±2% during each step of the simulation (data not shown).

(d) Simulations in a Solvated State. The simulations in a solvated state were carried out using the standard parameter files provided in X-PLOR for the TIP3p model of water. Duplication and translation of the standard water box provided in X-PLOR were used to generate a cubic water box capable of enclosing the anticodon-binding domain with an average layer of five water molecules around the protein surface. The domain was then enclosed in the water box using X-PLOR's solvating procedure. This procedure resulted in a molecular model containing 3492 atoms (1827 more than the model *in vacuo*). The initial model was then analyzed as described above, but with the following variations: the atomic charges of Asp, Glu, Arg, and Lys were not modified, and the initial energy minimization prior to simulated annealing was carried out for 10 000 cycles or until a gradient of 0.01 was reached to ensure proper equilibration of the solvated model.

RESULTS

MD Simulations of Wild-Type MetRS. The simulated molecular dynamics of the whole structure of MetRS was for the most part consistent with that of Ghosh et al. (1991; data not shown). The 30 amino acid helix−loop peptide (residues 439−470) is followed by a second helix (residues 471−491). Rapid motion of the loop peptide connecting these two helices is apparent. In wild-type MetRS, this motion affects only five residues of the loop, with a mean deviation of the α-carbons from the original coordinates (rms value) of about 4 Å after a 100 ps simulation. This motion is seen as a sharp peak when α-carbon rms values of motion are plotted along the sequence (Ghosh et al., 1991). A second, much slower motion affects the two helices. The two helices show a forward "shear-type" (Gerstein et al., 1994) motion with respect to the rest of surrounding secondary structure elements.

The possibility was explored of utilizing the anticodon-binding domain alone, instead of the whole enzyme, for the calculations. As described above, the domain−domain interactions were approximated by including in the calculation the residues at the C- and N-terminal ends of the anticodon-binding domain. Overall behavior similar to that seen in the whole structure was found for the domain by itself. The inclusion of a water box of 609 water molecules did not affect the quality of the observed motions, although a significant amount of solvent dampening was clear. The patterns of motions were similar for the whole structure of MetRS *in vacuo*, the anticodon-binding domain treated independently *in vacuo*, and the same domain when solvated (data not shown). Because successful simulations of domains within larger proteins have been reported previously, our results encouraged us to perform the rest of the dynamic simulations *in vacuo* with the anticodon-binding domain alone (Henry et al., 1985; Joseph et al., 1990; Ptaszek et al., 1994).

The simulations were applied to models of inactive and active representatives of a combinatorial library which was constructed as previously described. Table 1 shows sequence variants in the combinatorial library encompassing the region from K439 to G468. The helix extends to P460, which begins the loop. Required residues among the active variants include N452, R453, P460, W461, and K465. In addition to the wild-type protein, we selected five active (designated "AV") and five inactive variants ("IV") for our initial work. These particular variants were chosen because they represented many different kinds of active sequences including two with 17 alanine or serine residues (AV15 and AV15a), because the five inactive variants all accumulate as stable proteins *in vivo*, and because two of the highly substituted variants (AV15a and IV15b) switched between active and inactive phenotypes based on a single V463A substitution.

These sequences were built into models of MetRS. While the direct substitution into the crystal structure of residues of the highly substituted AV15 variant (Table 1) generated an unsuitable model, we found that an equilibrated, wild-type-like motion could be recovered through refinement of the models by simulated annealing.

Differences in Intramolecular Packing Energies for the Constructed Models. Calculations of the residue−residue interaction energies did not show any correlation between the sequence of a variant and its activity. The electrostatic energies of all models of mutated sequences (see below) showed a similar decrease of around 25% from the values obtained for wild-type MetRS, with no significant differences occurring between active or inactive variants. The van der Waals surface contact energies were more variable, with differences for variants of up to 10-fold from that for wild-type MetRS. However, no correlation could be found between a change in van der Waals contact energy and the active or inactive nature of any particular model (data and methods not shown). In fact, the van der Waals surface contact energy closest to wild-type MetRS was that for the inactive variant IV1 (a 2-fold increase in energy). For those reasons, we abandoned further attempts to relate experimental data to differences in intramolecular packing energies.

MD Simulations of Models Based on Sequences from the Combinatorial Library. All simulations were analyzed visually and compared by superimposing the rms values of motion (during 50 ps of dynamics simulation) for the α-carbons of each model. An apparent difference in dynamic behavior was found between all of the active and inactive models from the combinatorial library (described in more detail and explicitly illustrated below). In the case of the inactive models, the effect of the sequence changes resulted

Table 1: Combinatorial Library of MetRS[a]

	439																								468	Number of Ala/Ser
Active Variants																										
MetRS	K	A	V	R	E	I	M	A	L	A	D	L	A	**N**	**R**	Y	V	D	E	Q	A	**P**	**W**	V	V	A	**K**	Q	E	G	6
AV1	A	A	V	S	E	I	A	A	L	A	D	L	A	**N**	**R**	Y	V	S	E	Q	A	**P**	**W**	V	V	A	**K**	S	E	A	12
AV2	A	A	V	R	E	I	M	A	L	A	G	V	A	**N**	**R**	Y	V	S	E	Q	A	**P**	**W**	A	V	A	**K**	Q	V	A	10
AV3	A	A	V	R	E	I	M	A	L	A	S	L	A	**N**	**R**	Y	A	D	E	Q	S	**P**	**W**	A	V	A	**K**	S	E	A	12
AV4	A	A	V	R	E	V	V	A	A	A	D	L	A	**N**	**R**	Y	V	**N**	E	S	A	**P**	**W**	V	V	A	**K**	Q	E	A	10
AV5	A	A	V	R	E	I	V	A	L	A	D	V	A	**N**	**R**	A	V	G	E	S	A	**P**	**W**	V	V	A	**K**	Q	E	A	10
AV6	A	A	V	R	E	I	V	A	L	A	G	A	A	**N**	**R**	Y	V	D	E	S	A	**P**	**W**	A	V	A	**K**	S	E	A	12
AV7	A	A	V	S	E	V	A	A	L	A	G	L	A	**N**	**R**	Y	V	**N**	E	Q	A	**P**	**W**	V	A	A	**K**	Q	E	A	11
AV8	A	A	V	S	A	V	V	A	L	A	D	V	A	**N**	**R**	Y	V	D	E	Q	A	**P**	**W**	A	V	A	**K**	Q	E	A	11
AV9	A	A	V	S	A	T	V	A	L	A	D	V	A	**N**	**R**	Y	V	D	E	Q	A	**P**	**W**	V	A	A	**K**	Q	E	A	11
AV10	A	A	A	S	A	I	M	A	L	A	G	L	A	**N**	**R**	Y	V	S	E	Q	A	**P**	**W**	V	A	A	**K**	Q	E	A	13
AV11	A	A	A	R	E	V	M	A	L	A	N	A	A	**N**	**R**	Y	V	D	E	S	A	**P**	**W**	V	A	A	**K**	S	E	A	13
AV12	A	A	V	S	E	I	V	A	L	A	G	V	A	**N**	**R**	Y	V	D	E	S	S	**P**	**W**	A	V	A	**K**	Q	E	A	11
AV13	A	A	A	R	E	V	V	A	L	A	G	V	A	**N**	**R**	Y	V	**N**	E	S	S	**P**	**W**	V	V	A	**K**	Q	E	A	10
AV14	A	A	V	S	A	V	A	A	L	A	G	V	A	**N**	**R**	Y	V	**N**	E	Q	A	**P**	**W**	A	V	A	**K**	Q	E	A	12
AV15	A	A	A	S	A	I	M	A	L	A	S	A	A	**N**	**R**	Y	V	S	E	S	A	**P**	**W**	A	V	A	**K**	S	E	A	17
AV16	A	A	A	R	E	V	V	A	L	A	G	A	A	**N**	**R**	Y	A	D	E	Q	S	**P**	**W**	A	V	A	**K**	S	E	A	13
AV17	A	A	A	S	A	I	V	A	L	A	S	V	A	**N**	**R**	Y	V	S	E	Q	S	**P**	**W**	A	V	A	**K**	Q	E	A	14
AV18	A	A	V	S	A	V	T	A	L	A	G	V	A	**N**	**R**	Y	V	**N**	E	Q	S	**P**	**W**	V	A	A	**N**	S	E	A	12
AV15a	A	A	V	S	A	I	A	A	L	A	S	A	A	**N**	**R**	Y	V	S	E	S	S	**P**	**W**	A	V	A	**K**	S	E	A	17
	|——————————————**Helix**—————————————|																			|————**Loop**————|					
Inactive Variants																										
IV15b	A	A	V	S	A	I	A	A	L	A	S	A	A	**N**	**R**	Y	V	S	E	S	S	**P**	**W**	A	A	A	**K**	S	E	A	18
IV1	A	A	V	S	A	I	M	A	V	A	D	L	A	**N**	**R**	A	A	**N**	E	Q	S	**P**	**W**	V	V	A	**K**	S	E	A	13
IV2	A	A	V	R	E	V	V	A	L	A	D	V	A	**N**	**R**	A	V	G	E	**P**	A	**P**	**W**	V	A	A	**K**	S	E	A	11
IV3	A	A	V	R	E	V	M	A	V	A	S	L	A	**N**	**R**	A	V	S	E	S	S	**P**	**W**	V	A	A	**K**	Q	E	A	13
IV4	A	A	V	S	E	V	T	A	L	A	N	A	A	**N**	**R**	Y	V	G	E	**P**	A	**P**	**W**	A	A	A	**K**	S	E	A	13

[a] The table shows only those inactive sequences where five residues (shown in bold) essential for activity were retained. Six other residues (A440, A446, A448, E457, A451, and A463) were not changed, either because they were already alanine residues or because of their strong sequence conservation. The shaded positions mark residues different from those of the wild-type sequence. The sequences of variants AV1, AV3, AV15, AV15a, and AV15b were reported earlier, with the AV15a and AV15b variants designated as AV16 and IV16, respectively. Changes to residues other than Ala or Ser appeared in some instances for reasons for technical convenience (Kim et al., 1993c).

in an increase of the rms values between P460 and G480. The increase is limited to this region, and no other appreciable differences are observable in other parts of the models (data not shown). In particular, the essential (for tRNA recognition) P460, W461, and K465 display this perturbation in rms values. Mutants AV15a (active) and IV15b (inactive) made an especially striking comparison. Although the only difference between these two sequences is the substitution of V463 (AV15a) with an A (IV15b), the number of residues with rms values over 1.5 Å (between P460 and G480) increased from 5 in the AV15a model to 17 in the model of IV15b. The analysis of 10 previously reported mutant proteins, combined with that of 5 new mutants described below, enabled us to pinpoint the effect of residue motion on activity to the region of P460 to K465.

MD Simulations of Models Based on New Sequences Not Previously Tested Experimentally. We chose new mutants on the assumption that the structural effect revealed by dynamics simulations reflected changes in local packing interactions. We wanted to test the hypothesis that the enhanced mobility of the region containing P460−K465 was due to a general loss of packing interactions for the two helices which flank the mobile loop. We were also interested in obtaining new MD profiles that might allow us to find the most critical area perturbed by sequence changes.

We concentrated on two main interactions in the region mentioned above (Figure 1C,D). The first is the hydrophobic interaction between V463 and L473. This packing interaction is close to the loop (Figure 1C). It is also of particular interest because of the sensitivity of the activity and loop motion to the V463A substitution which distinguishes AV15a from IV15b. Two models with changes in the V463−L473 interaction were created in the context of the sequence of the wild-type enzyme. One contained the V463A replacement (V463A) and the other a V463A/L473A double replacement.

We also studied the perpendicular packing of Y454 with P415 (Figure 1D). The interaction of Y454 with P415 links the helix−loop peptide to a distal region in the sequence. To investigate the possible long-range effect of disrupting this particular interaction, we created two different models with a P415G substitution. One model was constructed with the mutation in the context of the wild-type sequence (P415G), while the second model contained the mutation in the context of AV16 (P415G_AV16). As a further study of the effects on packing of these two interactions in AV16, we also introduced P415G and L473A as a double replacement (P415G/L473A_AV16).

The analysis of the trajectories for these models showed that V463A, P415G, and P415G_AV16 variants have behavior similar to that of the wild-type MetRS and other active variants so that, based on fixation of motion, these three new variants were expected to be active.

As for the V463A/L473A double replacement, residue motion in the critical region of P460−K465 was similar to that of the wild-type and other active variants. In contrast, beyond K465 and up to G480, motions were much larger than seen in other active variants. We viewed the V463A/L473A double mutant as being particularly instructive for testing the idea that P460−K465 was actually the most

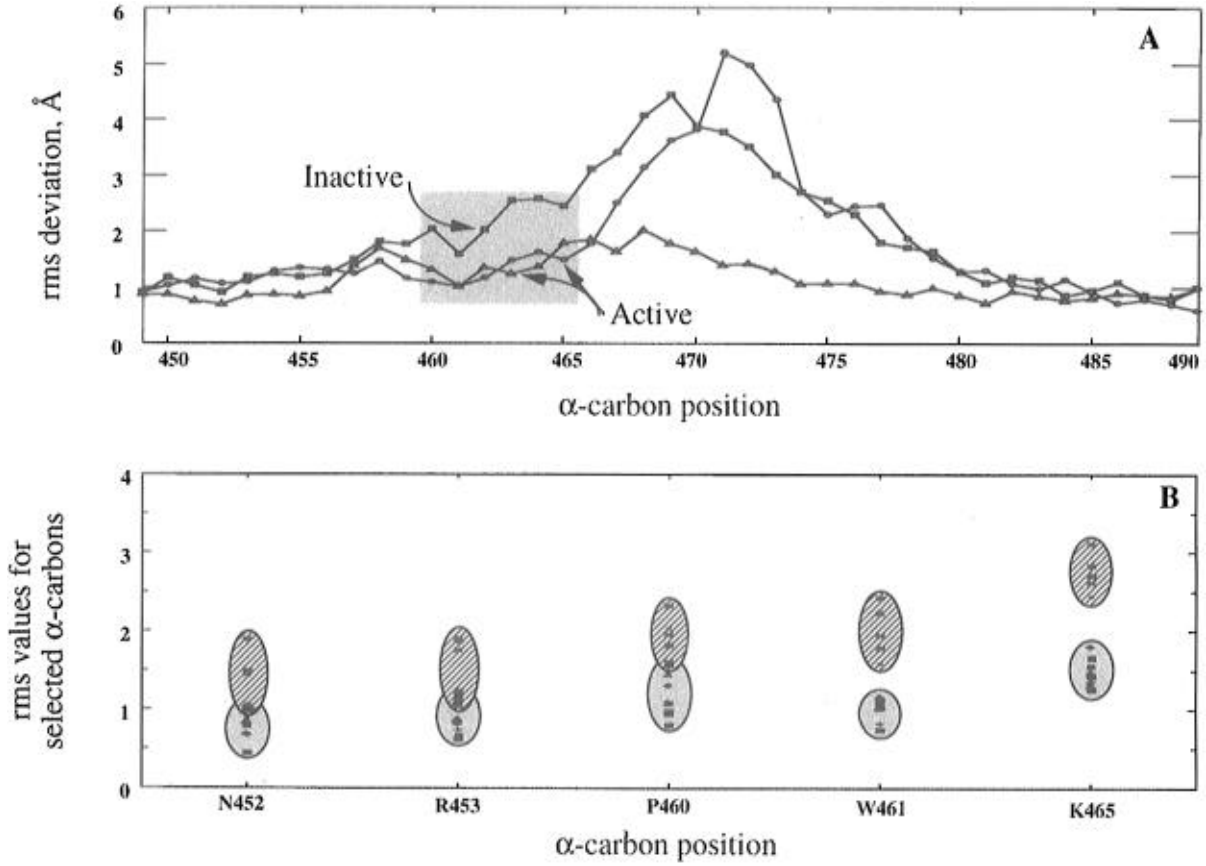

FIGURE 2: (A) Profile of rms α-carbon deviations for the 450−490 region from the 50 ps molecular dynamics simulations of three different MetRS models. The two models corresponding to active variants (V463A/L473A and AV15a) have a clearly different profile for the P460−K465 segment compared to the model for the inactive variant IV15b. The later variant differs from the active variant AV15a by a single V463A substitution (J = V463A/L473A; H = AV15a; B = IV15b). (B) Distribution of α-carbon rms values at five selected positions of 15 models of MetRS variants, including the wild-type enzyme. A segregation of models of active from inactive enzymes is clear at all positions. rms values for inactive variants are depicted with red symbols and those for active variants with green symbols. The standard error bars between values of active and inactive sequences (roughly depicted by the oval shapes) completely separate at W461 and K465.

critical region of the domain. Finally, the residue motions from P460 to G480 for the variant P415G/L473A_AV16 were similar to those observed for the rest of inactive variants.

Experimental Analysis of the New MetRS Variants. The five new MetRS mutants were created by site-directed mutagenesis, and the resulting enzymes were checked for complementation of a null (Δ*metG* MN9261/pRMS615) *metG* strain. This strain has a complete ablation of the chromosomally encoded gene for methionyl-tRNA synthetase, and it is supported by the pRMS615 maintenance plasmid containing the *metG* gene. The plasmid contains a temperature-sensitive replicon and is lost if the cells are incubated at 42 °C. When *metG* cells are transformed with a second plasmid containing a variant of MetRS and incubated at 42 °C, growth is observed only if the variant introduced has enough activity to replace the product of the wild-type *metG* that is lost with the maintenance plasmid. If a negative phenotype was observed, Western blot analysis of crude extracts (with anti-MetRS polyclonal antibodies) was used to determine whether the noncomplementing variant accumulated *in vivo*.

The four variants predicted to have limited motion of the P460−K465 region complemented the null strain (data not shown). Variant P415G/L473A_AV16, which displayed motion comparable to that seen with other inactive variants, failed to complement the *metG* null strain. However, extracts of cells expressing this variant had no detectable MetRS protein by Western blot analysis. Thus, this variant probably cannot fold into a stable structure.

Examples of Residue Motion Profiles in the Region L450 to Y490. As stated above, only the region between P460 and G480 was affected by the Ala and Ser substitutions placed in the 30 amino acid region between K439 and G468. In particular, the segment from P460 to K465 consistently segregated active from inactive variants, in the models of 15 mutants which were investigated by molecular dynamics simulations.

As examples of profiles obtained in the region from L450 to Y490, Figure 2A shows rms deviations of α-carbon atoms for two of the active enzymes (V463A/L473A and AV15a) and for one inactive variant (IV15b). The profile of residue motions for IV15b is similar to that seen with all other inactive variants, which typically have enhanced residue motions between P460 and G480. In contrast, active variants consistently have reduced motions in the P460−K465 region but have enhanced (V463A/L473A) or reduced (all other active models and wild-type MetRS) motions between positions 467 and 480. Particularly noteworthy is the dramatic change caused by the V463A substitution that distinguishes AV15a from IV15b (Figure 2A). These correlations of activity with limited motion of a few essential residues suggest that a depiction of the residue motions for

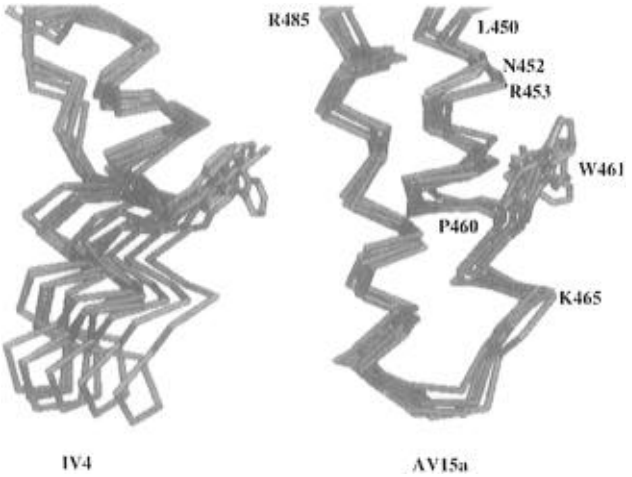

FIGURE 3: Comparison of the MD trajectories of a highly substituted active variant (AV15a, 17 Ala/Ser substitutions) and an inactive variant with less changes (IV4, 13 Ala/Ser substitutions), in the region from L450 to R485. Six different frames were extracted from each trajectory (at 0, 10, 20, 30, 40, and 50 ps), and the coordinates for all the α-carbons (red) and the side chain of W461 (green) are displayed to show the range of motion around the anticodon binding loop of each model.

a few selected positions in the sequence could be most revealing.

Comparison of Residue Motion for Active and Inactive Variants at Five Immutable Positions. We investigated the rms values for the α-carbons of the five essential residues (N452, R453, P460, W461, and K465) for all active and inactive variants (Figure 2B). Because they are essential for anticodon recognition, we felt that one or more of these five residues was most likely to show a difference (in α-carbon motion) between active and inactive residues. The general tendency of the α-carbon motions to segregate between active and inactive variants is clear for all five residues (Figure 2B). However, complete separation by residue motion of all 14 mutants is seen at W461 and K465. These are the only two residues which are out of the helix and within the loop. It is noteworthy that W461 has been proposed to interact with the "C" of the CAU anticodon of $tRNA^{Met}$ and was shown to be essential for discriminating that anticodon from the GAU anticodon of $tRNA^{Ile}$ (Auld & Schimmel, 1995; Ghosh et al., 1990).

Visualization of Atomic Motion in the Region Encompassing and Flanking the Immutable Residues. To provide a specific example of the differences in residue motion between an active and an inactive variant, we selected two highly substituted variants—one active and one inactive variant—and compared their dynamic trajectories by extracting their coordinates at 10 ps intervals. We focused on the region from L450 to R485, which encompasses the five immutable residues of the anticodon-binding helix−loop and the adjacent helix. The selected variants were AV15a and IV4, which have 17 and 13 Ala/Ser replacements, respectively. Thus, for this comparison, the active variant is more highly substituted than the inactive variant. For these two variants, the different trajectories for the α-carbon backbone and for the critical W461 side chain are evident (Figure 3). In particular, atoms of the less substituted IV4 variant show a far greater mobility. Similar comparisons with other variants show the same differences between active and inactive forms.

DISCUSSION

The simulations described here were for time frames of 50−100 ps. While longer simulations (1 ns) would be desirable, the computer time required with a large number of mutant proteins is excessive. The results of a 500 ps MD analysis of the same region (residues 370−510) in the wild-type protein showed that the rms deviations were relatively stable after 50 ps and remained so until the end of the simulation. The simulations also suggest that the effect of sequence variations in the helix−loop peptide on residue motions does not extend beyond a region of 40 or so amino acids (Figure 3). This observation is consistent with characterizations *in vitro* of inactive variants which showed that their catalytic sites in the N-terminal domain are undisturbed. For example, deletion of the entire region from Y454 to A464 in the anticodon-binding C-terminal domain yielded an inactive (for aminoacylation) mutant enzyme with kinetic parameters for amino acid activation that were essentially the same as those for the wild-type enzyme. Thus, loss of aminoacylation activity is directly caused by a deficiency in tRNA binding (Kim & Schimmel, 1992; Gale & Schimmel, 1995). The lack of coupling between catalytic domains of methionyl-tRNA synthetase has been seen in other studies as well. For example, the isolated C-terminal domain of MetRS has about the same affinity for the anticodon stem−loop hairpin of $tRNA^{Met}$ as does the whole enzyme (Gale & Schimmel, 1995). This observation shows that the N-terminal catalytic domain does not influence the conformation of the helix−loop peptide element in the C-terminal domain studied here.

This lack of coupling enabled us to use an *in vivo* complementation assay to assess tRNA recognition. For inactive (by complementation) variants, we generally have detected little or no aminoacylation activity *in vitro*. Because inactive mutant enzymes typically have robust adenylate synthesis capability, the complementation assay is a direct measure of the tRNA-dependent step of aminoacylation which, in this instance, is a measure of an interaction with the anticodon of $tRNA^{Met}$ (Kim et al., 1993c, 1994). In addition, *in vitro* kinetic analyses showed that, even for heavily substituted active variants, their aminoacylation activities are not greatly different than that observed for the wild-type enzyme (Kim et al., 1993c). This circumstance may result from an "all or none" effect (on motion-dependent activity) of the mutations placed in the helix−loop peptide. This effect, in turn, may reflect the strong sensitivity of the motion of multiple residues to a single change in the sequence. The comparison of AV15a and IV15b variants is a case in point, where a single V463A substitution increases from 5 to 17 the number of α-carbons with rms displacements greater than 1.5 Å.

All of the sequence variants of the helix−loop peptide studied in this work retained the five residues (N452, R453, P460, W461, and K465) which were found not replaceable in a large library of active mutants. Inactive variants that retain these five residues have exaggerated motions of the associated α-carbon atoms in the unbound protein. Assuming that the bound state places a severe constraint on the motion of these residues, the entropy cost for complex formation will be higher for those variants which have a greater motion of the essential residues in the unbound state. This cost is logarithmically related to the ratio of the

"effective" volume elements circumscribed by residue motions in the unbound and bound states. Given that motions of all five of the critical residues are sensitive to the sequence changes we studied, the entropy cost difference between active and inactive variants could easily amount to several kilocalories per mole (an exact calculation of the difference is not justified by the accuracy of the computational method employed and the approximations we have used). Thus, the need to fix residue motion in the unbound protein to reduce entropy could be a significant and general thermodynamic constraint on the sequences of RNA-binding peptide elements in proteins.

Spolar and Record (1994) presented a thermodynamic analysis of protein−DNA complex formation which emphasizes the role of induced fit. The energetic consequences of induced fit are manifested in large negative heat capacity changes and entropic deficits which accompany binding. From this perspective, our analysis suggests that inactive variants which retain essential residues can arise when the cost of induced fit (associated with a particular sequence) is too great and that, conversely, sequences which restrict residue motion in the free protein can reduce the entropic cost associated with induced fit.

ACKNOWLEDGMENT

We thank Drs. Carla Mattos, Carl Pabo, Tom Record, and Dagmar Ringe for comments and suggestions on the work and manuscript. We also thank Arturo Morales and Dr. Barry Henderson for helpful discussions.

REFERENCES

Auld, D. S., & Schimmel, P. (1995) *Science 267*, 1994−1996.
Biou, V., Yaremchuk, A., Tukalo, M., & Cusack, S. (1994) *Science 263*, 1404−1410.
Brünger, A. T., Kuriyan, J., & Karplus, M. (1987) *Science 235*, 458−460.
Brunie, S., Zelwer, C., & Risler, J.-L. (1990) *J. Mol. Biol. 216*, 411−424.
Burbaum, J. J., & Schimmel, P. (1991) *Biochemistry 30*, 319−324.
Burd, C. G., & Dreyfuss, G. (1994) *Science 265*, 615−621.
Gale, A., & Schimmel, P. (1995) *Biochemistry 34*, 8896−8903.
Gerstein, M., Lesk, A. M., & Shothia, C. (1994) *Biochemistry 33*, 6739−6749.
Ghosh, G., Pelka, H., & Schulman, L. H. (1990) *Biochemistry 29*, 2220−2225.
Ghosh, G., Kim, H. Y., Demaret, J.-P., Brunie, S., & Schulman, L. H. (1991) *Biochemistry 30*, 11767−11774.
Harrison, S. C. (1991) *Nature 353*, 715−719.
Henry, E. R., Levitt, M., & Eaton, W. A. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 2034−2038.
Joseph, D., Petsko, G. A., & Karplus, M. (1990) *Science 249*, 1425−1428.
Karplus, M., & Petsko, G. A. (1990) *Nature 347*, 631−639.
Kim, S., & Schimmel, P. (1992) *J. Biol. Chem. 267*, 15563−15567.
Kim, H. Y., Pelka, H., Brunie, S., & Schulman, L. H. (1993a) *Biochemistry 32*, 10506−10511.
Kim, S., Landro, J. A., Gale, A. J., & Schimmel, P. (1993b) *Biochemistry 32*, 13026−13031.
Kim, S., Ribas de Pouplana, L., & Schimmel, P. (1993c) *Proc. Natl. Acad. Sci. U.S.A. 90*, 10046−10050.
Kim, S., Ribas de Pouplana, L., & Schimmel, P. (1994) *Biochemistry 33*, 11040−11045.
Koshland, D. E., Jr. (1958) *Proc. Natl. Acad. Sci. U.S.A. 44*, 98.
Lim, W. A., Sauer, R. T., & Lander, A. D. (1991) *Methods Enzymol. 208*, 196−210.
Pabo, C. O., & Sauer, R. T. (1992) *Annu. Rev. Biochem. 61*, 1053−1095.
Ptaszek, L. M., Vijayakumar, S., Ravishanker, G., & Beveridge, D. L. (1994) *Biopolymers 34*, 1145−1153.
Rould, M. A., Perona, J. J., Söll, D., & Steitz, T. A. (1989) *Science 246*, 1135−1142.
Ruff, M., Krishnaswamy, S., Boeglin, M., Poterszman, A., Mitschler, A., Podjarny, A., Rees, B., Thierry, J. C., & Moras, D. (1991) *Science 252*, 1682−1689.
Ryckaert, J.-P., Ciccoti, G., & Berendsen, H. J. C. (1977) *J. Comput. Phys. 23*, 327−341.
Shepard, A., Shiba, K., & Schimmel, P. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 9964−9968.
Spolar, R. S., & Record, M. T. J. (1994) *Science 263*, 777−784.
Steitz, T. A. (1990) *Q. Rev. Biophys. 23*, 205−280.
Tan, R., Chen, L., Buettner, J. A., Hudson, D., & Frankel, A. D. (1993) *Cell 73*, 1031−1040.
Van Gunsteren, W. F., & Mark, A. E. (1992) *Eur. J. Biochem. 1992*, 947−961.
Vieira, J., & Messing, J. (1987) *Methods Enzymol. 153*, 3−11.
Vinson, C. R., Sigler, P. B., & McKnight, S. L. (1989) *Science 246*, 911−916.

BI960256A