

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/13600061>

# Determination of the amino acid requirements for a protein hinge in triosephosphate isomerase

ARTICLE *in* PROTEIN SCIENCE · JULY 1998

Impact Factor: 2.85 · DOI: 10.1002/pro.5560070702 · Source: PubMed

---

CITATIONS

29

---

READS

13

## 2 AUTHORS, INCLUDING:



Nicole Sampson

Stony Brook University

96 PUBLICATIONS 2,441 CITATIONS

[SEE PROFILE](#)

# Determination of the amino acid requirements for a protein hinge in triosephosphate isomerase

JEONGHOON SUN AND NICOLE S. SAMPSON

Department of Chemistry, State University of New York, Stony Brook, New York 11794-3400

(RECEIVED January 22, 1998; ACCEPTED March 25, 1998)

## Abstract

We have determined the sequence requirements for a protein hinge in triosephosphate isomerase. The codons encoding the hinge at the C-terminus of the active-site lid of triosephosphate isomerase were replaced with a genetic library of all possible 8,000 amino acid combinations. The most active of these 8,000 mutants were selected using *in vivo* complementation of a triosephosphate isomerase deficient strain of *E. coli*, DF502. Approximately 3% of the mutants complement DF502 with an activity that is above 70% of wild-type activity. The sequences of these hinge mutants reveal that the solutions to the hinge flexibility problem are varied. Moreover, these preferences are sequence dependent; that is, certain pairs occur frequently. They fall into six families of similar sequences. In addition to the hinge sequences expected on the basis of phylogenetic analysis, we selected three new families of 3-amino-acid hinges: X(A/S)(L/K/M), X(aromatic/β-branched)(L/K), and XP(S/N). The absence of these hinge families in the more than 60 known species of triosephosphate isomerase suggests that during evolution, not all of sequence space is sampled, perhaps because there is no neutral mutation pathway to access the other families.

**Keywords:** alpha/beta barrel; cassette mutagenesis; combinatorial library; hinge-bending; *in vivo* complementation; loop movement

Many enzymes have an intrinsic flexibility essential for their function (Kempner, 1993). Movement can occur on quite a large scale, and displacements of 10 Å are not unusual. Their degree of flexibility depends on the function of the enzyme and ranges from large domain movements to displacements of smaller segments. Comprised of contiguous segments of polypeptide, Ω loops, classified as nonregular secondary structure (Leszczynski & Rose, 1986; Fetrow, 1995), often participate in catalysis as flexible elements.

Ω Loops exist in a variety of conformations, have small end-to-end distances, and are typically comprised of 10–20 amino acids. They are situated on the surfaces of proteins, where large scale movements are free to occur. Catalytically important loops are a subset of Ω loops and displacements of 7 to 10 Å have been observed by comparison of X-ray crystal structures. The importance of these loops for proper catalytic function has been experimentally confirmed in many systems. They appear to act as flaps or lids for the active sites of enzymes. Their roles range from sequestering or protecting active intermediates from undesirable reactions to recruiting catalytic groups to active sites.

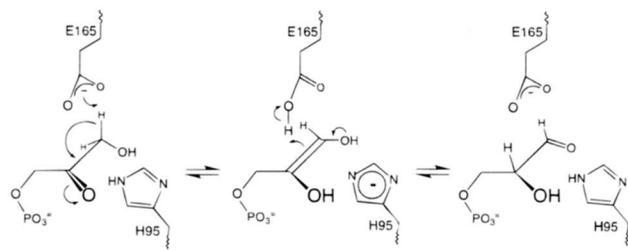
Loop movements can be categorized further into hinge motions and shear motions (Gerstein et al., 1994). (In addition, there are a

few other less defined motions.) These classifications are based on analysis of X-ray crystal structures of open and closed conformations of proteins. Protein hinges are characterized by the localization of motion to a few main-chain torsion angle changes. In contrast, shear motions are manifested as small side-chain torsion angle changes along an entire interface. The primary structural requirement for a protein hinge is that the residues in the hinge must not have tertiary structure packing constraints. In order to further define the structural requirements for a protein hinge, particularly in the case of loop motion, we have experimentally tested the amino acid sequence preferences for the C-terminal hinge of triosephosphate isomerase (TIM).

TIM is a glycolytic enzyme that catalyzes the interconversion of glyceraldehyde-3-phosphate (GAP) and dihydroxyacetone phosphate (DHAP) (Fig. 1). The reaction proceeds via an enediol(ate) intermediate. The kinetic mechanism of TIM has been extensively characterized (Knowles & Albery, 1977) and is well understood with respect to its structure (Knowles, 1991). Furthermore, the open and closed forms of the TIM loop and the differences between them have been analyzed in detail (Joseph et al., 1990; Derreumaux & Schlick, 1998).

Previously, it was demonstrated by mutagenesis that the active site lid of TIM sequesters the intermediate on the enzyme surface (Pompliano et al., 1990). Furthermore, the lid constrains the enediol(ate) in a planar conformation. When the lid is closed, protonation at C-1 or C-2 is favored over elimination of inorganic

Reprint requests to: Nicole S. Sampson, Department of Chemistry, State University of New York, Stony Brook, New York 11794-3400; e-mail: nicole.sampson@sunysb.edu.



**Fig. 1.** Reaction catalyzed by TIM.

phosphate to form methyl glyoxal. The lid open and closed conformations have been observed in X-ray crystal structures of the chicken, yeast, trypanosome, plasmodium, human, *Escherichia coli*, and *Bacillus stearothermophilus* enzymes (Lolis et al., 1990; Lolis & Petsko, 1990; Noble et al., 1991, 1993; Wierenga et al., 1991; Mande et al., 1994; Zhang et al., 1994; Delboni et al., 1995; Velankar et al., 1997). The closed loop structures are observed in the presence of inhibitors or sulfate, the open structures in the absence of ligand. The active site loop (loop 6) connects  $\beta$ -sheet 6 with  $\alpha$ -helix 6 (Fig. 2) and is on the opposite side of the  $\alpha/\beta$  barrel from the dimer interface. The loop motion is not affected by the dimer interface.

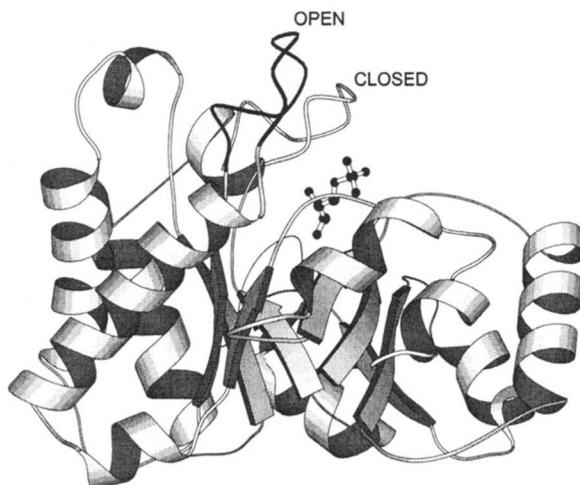
The 11-amino acid loop of TIM opens and closes as a rigid body, literally like a lid; the intervening loop residues do not change conformation. The lid pivots about two hinges that are comprised of three amino acids each (Joseph et al., 1990; Derreumaux & Schlick, 1998). The major changes are in the backbone dihedrals of residues 166–168 and 174–176 (Table 1). Consequently, residues 166–168 and 174–176 comprise two well-defined protein hinges. Across evolution, the sequences of the hinges are heavily, although not strictly, conserved with some variation between different species of TIM (Fig. 3). Since, by nature of their motion, protein hinges should not have tertiary structure packing

Organism	Accession number	N-terminal hinge	C-terminal hinge
<i>G. gallus</i>	P00940	Y E P V W A I G T G K T A T P	(23)
<i>S. pombe</i>	P07669	Y E P V W A I G T G K T G T P	(1)
<i>S. cerevisiae</i>	P00942	Y E P V W A I G T G L A A T P	(1)
<i>G. lamblia</i>	P36186	Y E P V W S I G T G V V A T P	(2)
<i>T. brucei</i>	P04789	Y E P V W A I G T G K V A T P	(20)
<i>B. stearothermophilus</i>	P00943	Y E P I W A I G T G K S S T P	(3)
<i>C. aurantiacus</i>	P96744	Y E P I W A I G T G D T A T P	(1)
<i>S. sp. (strain PCC 6803)</i>	Q59994	Y E P I W A I G T G D T C A A	(1)
<i>M. leprae</i>	P46711	Y E P V W A I G T G R V A S A	(1)
<i>E. coli</i>	P04790	Y E P V W A I G T G K S A T P	(3)
<i>V. marinus</i>	P50921	Y E P I W A I G T G K A A T A	(1)
<i>P. syringae</i>	P95576	Y E P V W A I G T G L T A S P	(1)
<i>X. flava</i>	P96190	Y E P I W A I G T G R T P T T	(1)
<i>B. aphidicola</i>	Q59179	Y E P I W S I G T G V S A D P	(1)
<i>Moraxella</i>	Q01893	Y E P V W A I G T G K V P T V	(1)
<i>M. jannaschii</i>	Q58923	Y E P P E L I G T G I P V S K	(1)
	*	***	170
			175
		active site loop	

**Fig. 3.** Alignment of amino acid sequences of TIM in the loop region using CLUSTAL W (Thompson et al., 1994). A representative subset illustrating the variation at the C-terminal hinge is presented, because there are 62 known sequences. The number of species containing each C-terminal hinge are indicated in parentheses. The sequences were obtained from the SWISS-PROT data bank (Bairoch & Boeckmann, 1994).

constraints (Gerstein et al., 1994), it is curious that more amino acid side-chain variability in the C-terminal hinge of TIM is not observed. The conservation of the N-terminal hinge may be due to the consideration that the active site base is glutamate-165 (Raines et al., 1986), i.e., adjacent to the hinge.

Later experiments suggested that the loop closes *after* the Michaelis complex is formed, rather than as it is formed (Sampson & Knowles, 1992). This observation is consistent with the paradigm for enzymatic catalysis: that stabilization of intermediates and transition states should be favored over stabilization of substrates and products. Furthermore, solid-state NMR experiments elegantly demonstrated that the rate of lid motion is on the same time scale as catalysis, and independent of the presence of substrate (Williams



**Fig. 2.** Ribbon diagram of TIM with the loop closed and glycerol 3-phosphate bound in the active site. The open loop (from the same crystal structure) is superimposed on the closed structure. This figure was created with MOLSCRIPT (Kraulis, 1991) using 6TIM (trypanosome) (Noble et al., 1991).

**Table 1.** Main chain dihedral angles for hinge residues of TIM<sup>a</sup>

Dihedral	Angle in closed lid <sup>b</sup> (deg)	Angle in open lid <sup>c</sup> (deg)
P166 $\phi$	-59	-78
$\varphi$	129	75
V167 $\phi$	-58	-58
$\varphi$	-43	-35
W168 $\phi$	-59	-54
$\varphi$	-19	-27
K174 $\phi$	-90	-74
$\varphi$	75	164
T175 $\phi$	-62	-121
$\varphi$	146	127
A176 $\phi$	-71	-92
$\varphi$	153	133

<sup>a</sup>For all residues,  $\omega$  ranged from 173° to 180°.

<sup>b</sup>Closed loop measurements were taken from 1TPH (chicken) with phosphoglycolohydroxamate (PGH) bound and are from the A subunit (Zhang et al., 1994).

<sup>c</sup>Open loop measurements were taken from 6TIM (trypanosome) and are from the A subunit (Noble et al., 1991). The B subunit in 6TIM is closed and has very similar  $\phi$  and  $\varphi$  values as 1TPH.

& McDermott, 1995). Dynamics simulations have confirmed this time scale (Derreumaux & Schlick, 1998). Thus, lid motion is not ligand induced, but rather has evolved to occur slowly enough that deprotonation and reprotonation occur before the lid opens again. The closure of the lid is primarily stabilized by formation of an intramolecular hydrogen bond between the hydroxyl of tyrosine-208 and the amide-NH of alanine-176 (Sampson & Knowles, 1992; Derreumaux & Schlick, 1998). These data raised the question: what constitutes a protein hinge that opens and closes with the proper time constant?

The determinants of the time scale for motion have not been experimentally analyzed. Thus, it was unclear whether backbone torsional constraints, solvation of side chains, or even subtle tertiary packing interactions would govern loop motion and hinge function. Thus, we undertook an investigation of the structural requirements for a protein hinge in TIM and analyzed the C-terminal hinge of TIM, i.e., residues 174–176. Because the interplay of the various factors potentially affecting hinge function was unclear, we chose to construct a combinatorial library containing all possible three amino acid hinges at positions 174–176. This hinge was chosen over the N-terminal hinge at residues 166–168 because of its distance from the active site base glutamate-165. We wanted to isolate our study as much as possible to the role of the hinge. We screened the members of this library for catalytically active members using *in vivo* selection. The functional hinges were identified by sequencing the hinge region of the gene. Our experiments were designed to test the effects of neighboring residues in a hinge on each other, i.e., to determine if different 3-amino-acid sequences may function as a hinge, and if so, what sequences.

## Results

### Construction of hinge library pTMCA

A plasmid was constructed that produced reduced amounts of TIM to allow *in vivo* selection. A cassette encoding the hinge library was synthesized enzymatically from an oligonucleotide, in which each of the three hinge codons was replaced with NNS; the remainder of the oligonucleotide corresponded to the wild-type TIM gene sequence. This cassette was inserted into pTM04 to yield pTMCA, a library of hinge mutants. Ligation of the cassette deleted the *PmlI* site present in pTM04 and allowed for a rapid screen of subcloning efficiency. Using Klenow DNA polymerase to synthesize the second strand ensured that each cassette was perfectly complementary, and that all 32,768 possible base combinations would be present in the library. Use of S (1:1, C:G) at the third position of each hinge codon reduced the variation in amino acid frequency, due to coding redundancy and eliminated two stop codons. The efficiency of construction was verified by sequencing randomly selected mutants; the expected base usage at the hinge codons was observed. After translation, our library contains 9,261 combinations from 20 amino acids and the amber codon.

### Classification of active C-terminal hinge mutants

The hinge library was subjected to selection by *in vivo* complementation of DF502 (*tim*<sup>-</sup>). Those colonies that grew were isolated and the C-terminal hinge sequence determined. Initially, all hinge mutants that complemented DF502 were sequenced. After it became apparent that the number of mutants able to complement

DF502 was quite large, we screened the specific activity of crude cell lysates and sequenced only those mutants that had greater than 70% of wild-type activity. As shown in Tables 2 and 3, we sequenced 51 active (>70% wild-type activity) hinge mutants, 85 semi-active (10–70% wild-type activity), and 44 slightly-active (<10% wild-type activity). We used numerical simulation of the selection process to estimate the total number of mutants in the active category. Based on the number of active mutants selected once and the number of active mutants selected twice, we could simulate selection from a pool of active mutants using a random number generator. We adjusted the size of the active mutant pool in the simulation until the number of duplicates chosen in the simulation matched the number we had selected. From this simulation, the total number of active mutants is approximately 200. Thus, we have identified about 25% of the active mutants. Although it was possible to sequence more mutants, the emergence of patterns in our sequenced hinges convinced us that we had sufficient sequence information for the analysis described below.

We chose to define those mutants having a specific activity of 70% or greater relative to wild type as active. Although it may seem an arbitrary value, 70% represented the best compromise between three considerations. First, we were only interested in those mutants that were essentially wild type in activity. Second, the specific activity measurements control for variation in growth

**Table 2.** Sequences of semi-active and slightly-active selected mutants

Semi-active (10–70%)			Slightly-active (<10%)	
ATL	KPA	NWS	APL	KNN
CIL	KPN	QCL	CCH	KWD
CPM	KPT	QMA	CMR	LAF
CRA	KSE	QSI	CTC	NEG
CTS	KSG	RCS	CWV	NKM
CVL	KST	RKT	DAF	NSV
DAL	KTS	RLP	DFR	QNS
DIL	KYK	RTL	DRT	RWG
DLL	LFA	SLL	FEA	SCP
DNL	MTS	SSL	GVM	SKT
DPI	NAH	SST	GWC	SLI
FPC	NCG	SYT	GYQ	SNM
HGK	NCL	TFY	HCF	TEL
HPS	NFL	TLP	HLC	TIP
HSM	NFS	TSA	HNI	YSV
HST	NFT	VAY	HPF	YTI
HYL	NIG	VWA	HSH	*PG
ICL	NIL	VWP	IRY	*TS
IIK	NMP	WSS	IVL	L*L
KAA	NPQ	YAS	IWP	Q*S
KAC	NSA	VCA	KCL	NP*
KAP	NSC	YLP	KIT	YC*
KAS	NSH	YSA		
KAY	NST	YSK		
KGL	NTH	YTL		
KHL	NTL	YYA		
KIL	NTP	II*		
KML	NVA			
KMP	NYS			

<sup>a</sup>Asterisk represents the amber codon.

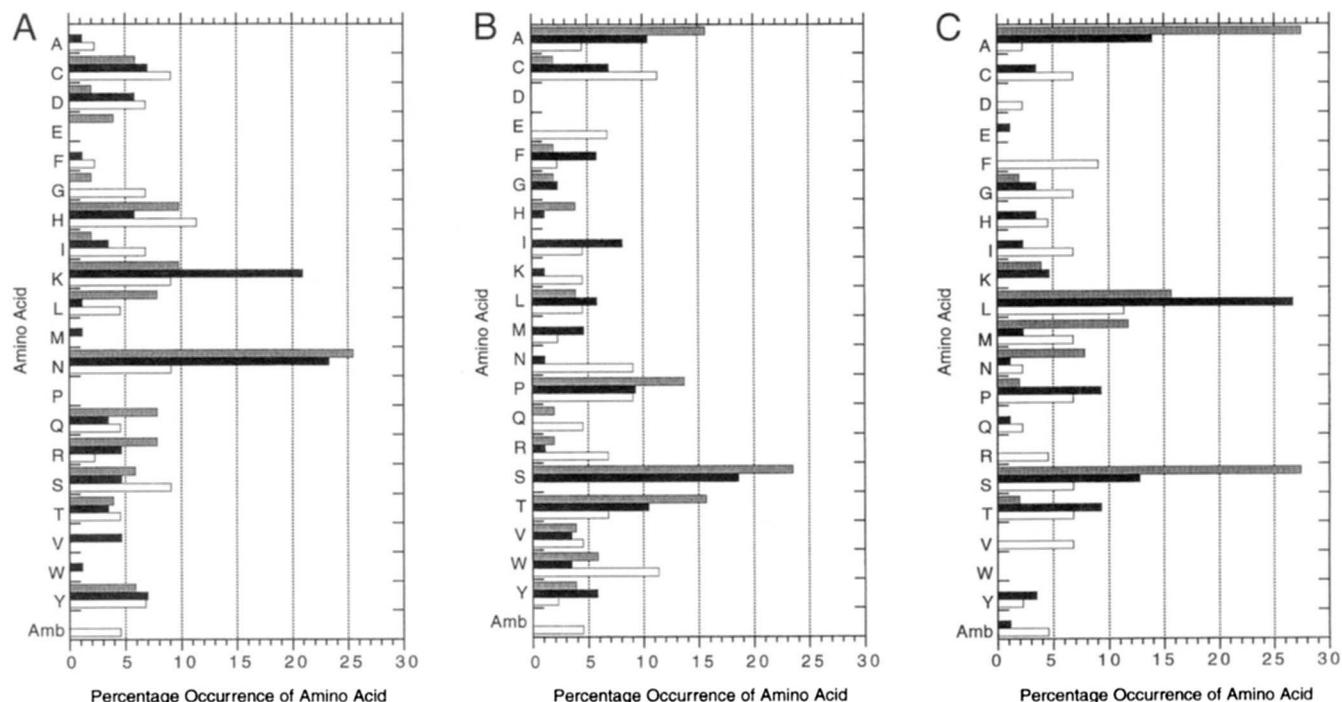
**Table 3.** Active mutants arranged by family<sup>a</sup>

1. XSA, XSS, XAS			2. X(Φ/β)(A/S) <sup>b</sup>		
GSA (100)	CSS (83)	CAS (182)	LWA (108)	NLS (76)	NTS (180)
HSA (87)	DSS (134)	NAS (105)	NFA (106)	SLA (92)	NVS (112)
LSA (190)	HSS (85)	NAN (76)	NHA (100)	EYA (182)	HTS (170)
	KSS (109)		TWA (77)	KRA (73)	QTA (86)
	NSS (135)		YWA (106)	NTN (112)	RTA (76)
3. X(A/S)(L/K/M)			4. X(Φ/β)(L/K)		
LAK (171)	KAM (73)	ICM (72)	KTK (82)	SYL (94)	LHL (82)
QL (76)	NAM (78)	RQL (80)	QTL (110)	KVL (92)	
YAL (123)	NSM (119)				
YSL (100)	RAM (84)				
5. XP(S/N)			6. Unclassified		
CPS (91)	HPN (113)	NPM (169)	EGA (87)	TSG (71)	RTP (83)
QPS (70)	NPN (79)	HPT (121)			
SPS (167)					

<sup>a</sup>Percent specific activity relative to wild-type TIM in parentheses.<sup>b</sup>β, β-branched amino acids; Φ, aromatic amino acids.

rates during cell culture; however, some mutants may be expressed at slightly higher levels than others, introducing error into the specific activity measurement. Third, the number of clones that must be sequenced to identify all active mutants is quite large. Seventy percent activity ensured that we had sequenced a signif-

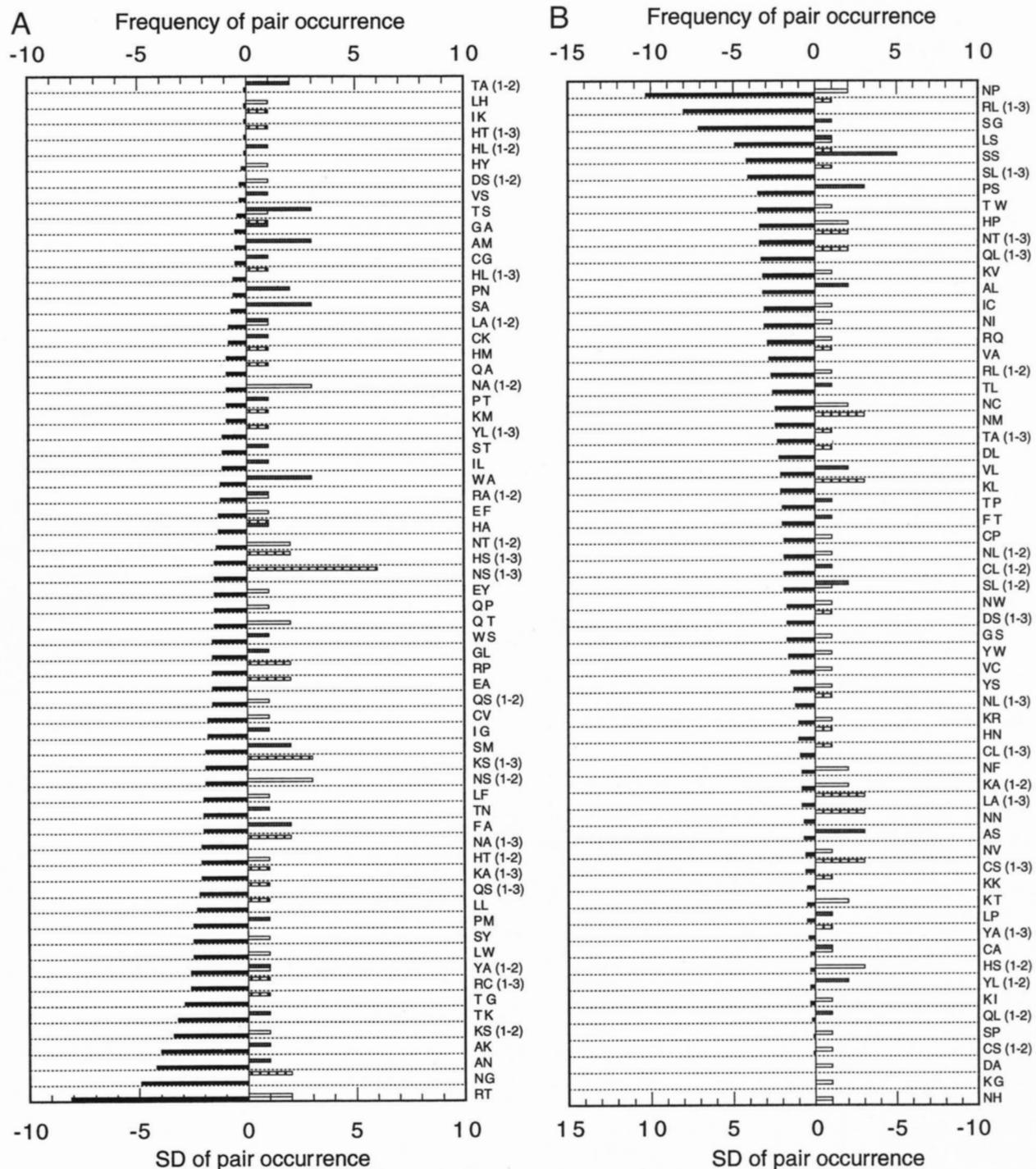
icant fraction of the active mutants. Finally, we note that our *in vivo* selection selects not only for the most catalytically active mutants, but also those proteins most stable at 37°C. If mutant proteins are unstable and degraded rapidly, the cell lysate specific activity will be low and the mutants remain unselected.

**Fig. 4.** Bar graphs of percent amino acid composition of hinges at each position for each subset of activity. Gray bar, active; black bar, semi-active; white bar, slightly active. **A:** Position 174. **B:** Position 175. **C:** Position 176.

## Analysis of hinge mutants

After classification of the mutants into three categories by specific activity, we examined the amino acid usage at each position

(Fig. 4). Comparison of each category of activity reveals that we are, in fact, performing a selection. Specific amino acids are preferred at each position of the C-terminal protein hinge. The activity of the mutants in the semi-active and slightly-active categories



**Fig. 5.** Bar graphs comparing the probability of amino acid pairs occurring in all proteins in standard deviation units (black bars) (Cserzö & Simon, 1989) and the frequency of occurrence for the active mutant category in positions 174–175 (white bars), 175–176 (gray bars), and 174–176 (striped bars). Numbers in parentheses on the x-axis indicate whether the pair occurs as  $n$  and  $n + 1$ , or  $n$  and  $n + 2$  amino acids in the protein sequence. **A:** Pairs of amino acids that have a negative standard deviation of occurrence. **B:** Pairs of amino acids that have a positive standard deviation of occurrence. Note that the y-axis for SD of pair occurrence has been inverted compared to A for clarity.

containing amber codons are a result of amber suppression. These mutant constructs express full-length protein (data not shown).

Our experiments were designed to assess the interplay of the amino acids at each position, and how they contribute to a hinge. We next examined, therefore, the occurrence of pairs of amino acids in the active category. Cserzö and Simon (1989) have analyzed the occurrence of amino acid pairs in a database of proteins containing over a million amino acids. They analyzed nearest neighbor pairs ( $n, n + 1$ ) out to the ninth position ( $n + 9$ ) without regard for type of secondary structure. The frequency of occurrence was then normalized for amino acid usage and represented as a probability in standard deviation units. A high positive standard deviation indicates that a pair was very likely to occur (greater than random frequency). A very negative standard deviation indicates that a pair occurs less frequently than random chance would dictate. In Figure 5, we compare the frequency of occurrence of amino acid pairs in the active mutant category, to the expectation of occurrence in any protein. For all three pairs (174–175, 175–176, 174–176) in the protein hinge, we see that statistically unfavored pairs are present at the same frequency as those favored. Thus, the C-terminal hinge of TIM utilizes amino acid combinations that are not commonly found in proteins. They may be particular to a protein hinge.

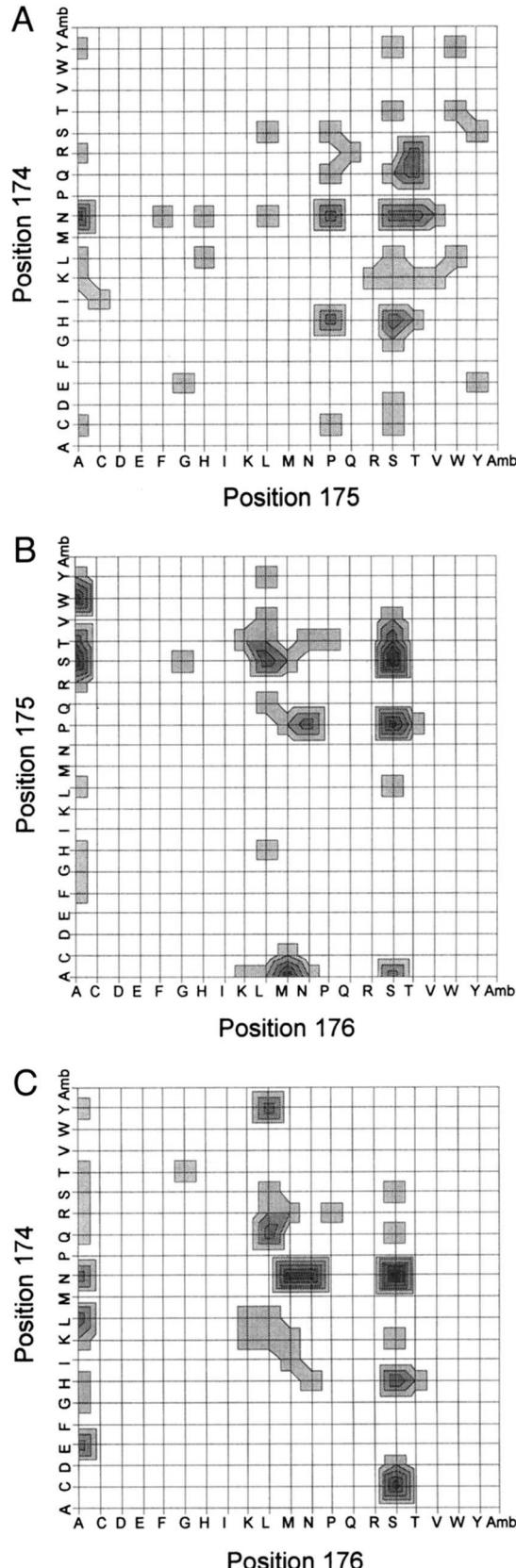
In order to assess the pairs that occur, and to classify them into related families of sequence, we prepared contour plots of the frequency data (Fig. 6). The plots highlight the narrow amino acid preference at position 176. The position 175–position 176 pairs (Fig. 6B) were organized into six families of mutants (Table 3).

## Discussion

In our investigation of the amino acid requirements of the C-terminal hinge in TIM, we found that the second and third positions (175 and 176) had the most stringent requirements. At position 174 (Fig. 4A), which is lysine in chicken TIM, we observe that three amino acids—asparagine, lysine and sometimes histidine—are preferred. Aromatic,  $\beta$ -branched, acidic, and small amino acid side chains are not present, except for tyrosine, in the active category. In addition, proline, the most conformationally restricted amino acid, is absent. The prevalence of asparagine over lysine was unexpected, because it is not observed in wild-type TIM from over 60 species. This prevalence is not due to codon bias; lysine and asparagine are each encoded by one codon in our library.

At position 175 (Fig. 4B), which is threonine in chicken TIM, we observe that 35% of the mutants contain serine and alanine; threonine is the third most frequently occurring amino acid. These amino acids are all reasonable with respect to wild type from many species. Amino acids with carboxylic acid or amide side-chain functionality are absent. The unexpected result was that proline occurred at the same frequency as threonine.

The values of  $\phi$  and  $\psi$  at position 175 in the closed form are  $-62^\circ$  and  $146^\circ$  (Table 1). These values are just outside region 4 in a Ramachandran plot, the only region permitted for valine and isoleucine (Nemethy et al., 1966). Technically, these  $\phi, \psi$  angles do not represent an unfavorable conformation for threonine; however, a solvated threonine is isosteric with valine. Furthermore, the second most frequent amino acid in wild-type sequences at position 175 is valine, e.g., in trypanosome TIM (Swinkels et al., 1986). In the closed form, valine has a  $\phi$  value of  $-57^\circ$  (Noble et al., 1991), confirming the unusual conformation of this residue in the fully closed TIM. Consideration of these conformational



**Fig. 6.** Contour plot of amino acid pairings of active mutants. The contour levels represent frequency of occurrence. Contours range from 0–8% occurrence in the active category in 1% increments (light to dark). **A:** Positions 174–175. **B:** Positions 175–176. **C:** Positions 174–176.

preferences for the closed lid elucidates why we select serine, alanine, and proline. Serine and alanine, are of course, much less conformationally restricted. The five-membered ring of proline seriously limits rotation about the N-C $\alpha$  bond, and the value of  $\phi$  is limited to approximately  $-60^\circ$  (Leach et al., 1966). The closed lid may be stabilized by the substitution of proline at this position. The impact on the stability of the open lid and the dynamics of movement will be the subject of future investigation.

At position 176 (Fig. 4C), which is alanine in chicken TIM, we see the most stringent amino acid requirements. Acidic, aromatic,  $\beta$ -branched, and basic amino acid side chains are absent. Fifty percent of the active mutants selected have alanine or serine. These two amino acids also predominate in wild-type TIMs from different species. Twenty percent of the active mutants have leucine at position 176. In the wild-type chicken structure, alanine-176C $\alpha$  is projecting away from solvent, toward the N-terminal hinge, in both the open and closed forms, and the C $\beta$  is packed inside the lid (Fig. 7). There is actually a small cavity below the C $\beta$  formed by tyrosine-164C $\epsilon$ , glutamine-180C $\beta$ , alanine-181C $\beta$ , tyrosine-208C $\epsilon$ ,

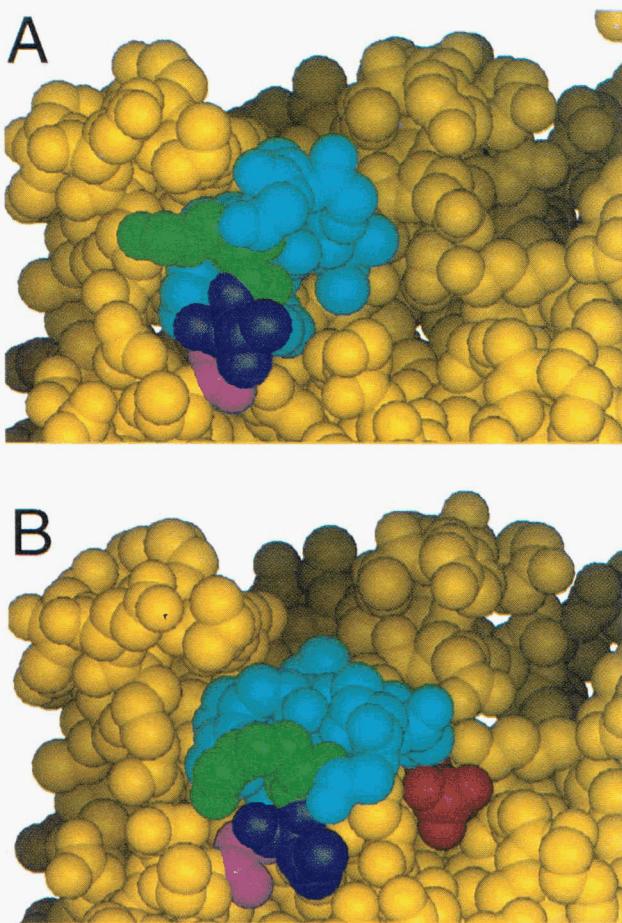
histidine-185C $\epsilon$ , and valine-184C $\beta$ . In both structures, this cavity is shielded from solvent by glutamine-180. It is possible that the straight-chain aliphatic amino acids, e.g., leucine, lysine, and methionine, pack into this cavity providing additional stabilization of the loop by burying hydrophobic surface. Selection of these active mutants suggests that tertiary packing interactions can contribute to hinge function. Exactly how these buried hydrophobic groups contribute to the dynamics of the loop is not clear.

The most noticeably absent amino acid at all three positions is glycine. *A priori*, one might predict that protein hinges should contain a high percentage of glycine, because this amino acid is the most conformationally versatile. The low occurrence of glycine (3 out of 51 active mutants) and the absence of any mutants with more than one glycine highlight the importance of controlled flexibility at a protein hinge. The active site lid must open and close. If motion is too rapid, however, the function of the lid will be lost; enediol(ate) intermediate will be released from the active site, and catalytic efficiency reduced.

The active mutants were categorized into families using the pairs that were selected at positions 175 and 176 (Fig. 6; Table 3). Similar amino acids at each position were grouped into the same family, e.g., serine and alanine. Families 1 and 2 are essentially wild-type sequences. Families 3 and 4 have what can be considered large aliphatic hydrophobic side chains at position 176. In family 3, position 176 is paired with small amino acids, alanine, and serine but in family 4 is paired with  $\beta$ -branched ( $\beta$ ) or aromatic ( $\Phi$ ) amino acids. Family 5 is most interesting. A proline is at position 175. Further structural analysis is required to determine if it is *cis* or *trans*, or if isomerization is occurring. The sixth category contains three unclassified mutants. The RTP mutant is actually a wild-type hinge from *X. flava*s and very similar to the KVP hinge from *Moraxella*. The EGA and TSG mutants could be considered part of family 1. We have not classed them in family 1 to highlight that in rare sequence contexts, glycine can make a suitable hinge. In summary, we have determined that there is more than one type of sequence that will serve as a protein hinge at the junction between a lid and an  $\alpha$ -helix. These families will be the subject of future investigations to determine the relationships between hinge sequences and kinetic profiles and structure.

### Conclusions

Examination of wild-type sequences from over 60 different species of TIM suggests that there are very specific requirements for protein hinges, because there is a very small amount of sequence drift. Using a combinatorial library containing all possible three amino acid hinges, we have screened for those combinations that form a catalytically active protein. We conclude that, in fact, there are amino acid preferences for a protein hinge. Moreover, these preferences are sequence dependent; that is, certain pairs occur frequently. Interestingly, in addition to the hinge sequences expected on the basis of phylogenetic analysis, we selected three new families of 3-amino-acid hinges: X(A/S)(L/K/M), X( $\Phi/\beta$ )(L/K), and XP(S/N). The selection of these families suggests that tertiary packing interactions may be more important for hinge function than previously realized. The absence of these hinge families in the 60 known species of TIM implies that there are other evolutionary criteria for which we do not select in our *in vivo* selection system. Alternatively, the codons for the hinges may be restricted to families 1 and 2 in nature because there is no neutral mutation pathway to access the other families (Gloss et al., 1996).



**Fig. 7.** CPK model of C-terminal hinge of TIM using the 6TIM structure (trypanosome) (Noble et al., 1991). Green, position 174 (lysine-176 in structure, threonine-175 in chicken TIM); blue, position 175 (valine-177 in structure, alanine-176 in chicken TIM); magenta, position 176 (alanine-178 in structure, alanine-176 in chicken TIM). The remainder of the loop is shown in cyan. **A:** Open loop (A subunit). **B:** Closed loop (B subunit with bound glycerol 3-phosphate, red).

## Materials and methods

### Materials

All commercial chemicals were used as obtained without further purification, and all solvents were dried and distilled by standard methods prior to use. NADH, DL-glyceraldehyde 3-phosphate (diethyl acetal, monobarium salt), and EDTA were purchased from Sigma Chemical Co. (St. Louis, Missouri). Oligonucleotide primers were ordered from Integrated DNA Technologies Inc. (Coralville, California) and are listed in Table 4. For site-directed mutagenesis, the Quick Change™ site-directed mutagenesis kit from Stratagene (La Jolla, California) was used. For DNA sequencing, the Sequenase Version 2.0 DNA sequencing kit from United States Biochemical (Cleveland, Ohio), the CircumVent™ thermal cycle dideoxy DNA sequencing kit with Vent<sub>R</sub> (exo<sup>-</sup>) DNA polymerase from New England Biolabs (Beverly, Massachusetts), or the ABI PRISM™ dye terminator cycle sequencing kit with AmpliTaq DNA PolymeraseFS from Perkin Elmer (Foster City, California) was used according to the manufacturer's instructions. Deoxyadenosine 5'-α-[<sup>35</sup>S]-thiotriphosphate and deoxyadenosine 5'-γ-[<sup>33</sup>P]-triphosphate triethylammonium salts were purchased from Amersham Life Science (Arlington Heights, Illinois). Restriction endonuclease EcoRI was purchased from Boehringer-Mannheim Biochemicals (Indianapolis, Indiana), and *Sca*I, *Pst*I, *Pml*I, *Age*I, Klenow fragment, and T4 DNA ligase were from New England Biolabs (Beverly, Massachusetts). Alkaline phosphatase was purchased from United States Biochemical (Cleveland, Ohio). All other reagents were purchased from Fisher Scientific, Inc. (Springfield, New Jersey).

The components of reaction buffers used for restriction digestion, ligation, phosphorylation, polymerization, and protein purification and assay are listed below. Buffer A: 100 mM triethanolamine-HCl, 1 mM EDTA pH 7.9 (at 30 °C); Buffer B: 10 mM bis tris propane-HCl, 10 mM MgCl<sub>2</sub>, 1 mM DTT pH 7.0 (at 25 °C); Buffer C: 10 mM tris-HCl, 10 mM MgCl<sub>2</sub>, 50 mM NaCl, 1 mM DTT, 25 µg/mL BSA pH 7.9 (at 25 °C); Buffer D: 10 mM Tris-HCl buffer pH 7.8 (at 25 °C), Buffer E: 100 mM triethanolamine-HCl, 10 mM EDTA pH 7.8 (at 25 °C). Rich medium was Luria broth (10 g tryptone, 5 g yeast extract, 5 g NaCl per liter). Minimal medium was M63 salts containing CAS amino acids (0.5%, w/v), glycerol (0.2%, w/v), FeSO<sub>4</sub> (0.5 mg/L), thiamine (1 mg/L), L-histidine (80 mg/L), streptomycin (100 mg/L), and ampicillin (200 mg/L).

Bromohydroxyacetone phosphate (BHAP) was prepared as described by de la Mare et al. (1972). Glycerol-3-phosphate

dehydrogenase was obtained from Boehringer-Mannheim (Indianapolis, Indiana) and residual TIM activity was removed by treatment with BHAP for 1 h; followed by ultrafiltration into buffer A.

*E. coli* strains XL1Blue and DH5 $\alpha$  were used for plasmid construction. *E. coli* strain DF502 (strept<sup>R</sup>, tpi<sup>-</sup>, and his<sup>-</sup>) used for in vivo selection was a generous gift from Drs. D. Fraenkel and J.R. Knowles and was described previously (Straus & Gilbert, 1985). The plasmids pBSX1cTIM (wild type) (Hermes et al., 1989), and pKKTIM (H95N) (Komives et al., 1991) were generous gifts from Drs. E. Komives and J.R. Knowles. pKK223-3 was purchased from Pharmacia (Piscataway, New Jersey).

### Plasmid construction

Methods for plasmid preparation and propagation were standard unless otherwise described (Sambrook et al., 1989).

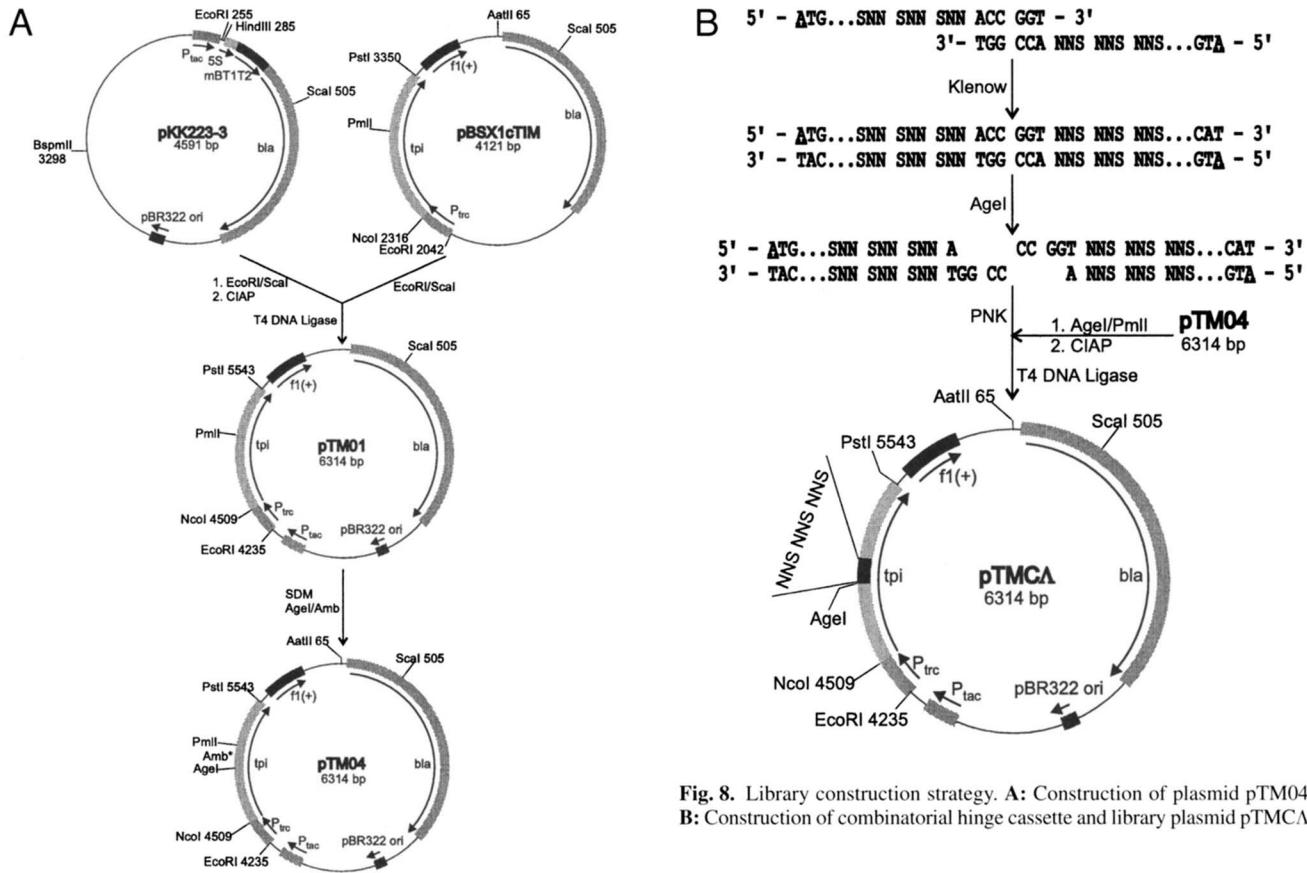
A plasmid was constructed that expressed wild-type chicken isomerase in *E. coli* at reduced levels (7 mg/L) compared to the pBSX1cTIM expression vector (115 mg/L) (Hermes et al., 1989). pTM01 was constructed by subcloning the 2.6 kb *Eco*RI to *Scal* fragment from pBSX1c TIM (wild-type chicken TIM) into the 3.7 kb *Eco*RI to *Scal* fragment of pKK223-3 (Fig. 8A). The plasmids were restricted and ligated to generate a 6.3 kb plasmid carrying the ampicillin resistance gene (*bla*) and containing the chicken TIM gene (*tpi*) behind the *trc* promoter. This plasmid is similar but not identical to the ptsTIM plasmid previously described (Hermes et al., 1990). The identity of the construct was verified by restriction digestion and analysis of the fragments by agarose gel electrophoresis. The ability of pTM01 to complement DF502, a strain of *E. coli* in which the triosephosphate isomerase gene has been deleted (*tpi*<sup>-</sup>) (Fraenkel, 1986), was verified by plating DF502(pTM01) transformed cells on minimal medium containing 0.2% glycerol. After 30 h at 37 °C, the colonies were 2.5–3 mm in diameter.

pTM02 was constructed in a similar fashion. The 1.3 kb *Pst*I to *Scal* fragment from pBSX1cTIM was subcloned into the 5.0 kb *Scal* to *Pst*I from pkkTIM (H95N chicken TIM). This provided a plasmid identical to pTM01 except that the TIM gene has the point mutation H95N. pTM02 does not complement DF502.

The nucleotide sequence of pTM01 was analyzed with GMAP (Raghava & Sahni, 1994) to find potential unique restriction sites that could be introduced without altering the amino acid sequence of the protein. pTM03 was constructed by site-directed mutagen-

**Table 4.** Primers used for library construction

Primer #	Purpose	Plasmid constructed	Primer
1	Introduction of <i>Age</i> I site anti-coding sense	pTM03	5'-GTA GCA GTT TTA CCG GTT CCG ATA GCC CAA ACT GG-3'
2	Introduction of <i>Age</i> I site coding sense	pTM03	5'-CCA GTT TGG GCT ATC GGA ACC GGT AAA ACT GCT AC-3'
3	Introduction of amber codon anti-coding sense	pTM04	5'-CCT CCT GAG CCT ATT GGG GAG TAG CAG-3'
4	Introduction of amber codon coding sense	pTM04	5'-CT GCT ACT CCC CAA TAG GCT CAG GAG G-3'
5	C-terminal hinge cassette self-complementary anti-coding sense	pTMCA	5'-ATG GCT TTT GAG CCA GCC TCT CAG CTT CTC ATG AAC CTC CTG AGC CTG TTG GGG AGT SNN SNN ACC GGT-3'



**Fig. 8.** Library construction strategy. **A:** Construction of plasmid pTM04. **B:** Construction of combinatorial hinge cassette and library plasmid pTMCA.

esis of pTM01. An *Age*I restriction site was introduced at bp 516 of the TIM gene, thus changing codon 172 from ACT (thr) to ACC (thr) on the 5'-side of the codons for the C-terminal hinge residues. The mutation was introduced by PCR mutagenesis using the Quick Change™ Mutagenesis Kit according to the manufacturer's instructions. Primers 1 and 2 (Table 4) were used. The mutation was verified by restriction digestion and sequencing of the TIM gene. A *Pm*I site was present on the 3'-side of the C-terminal hinge. These restriction sites were unique in the plasmid.

In addition, an amber codon was introduced into the cassette region to prevent in vivo selection of wild type (Fig. 9). pTM04 was constructed from pTM03 in a fashion analogous to the construction of pTM03. Primers 3 and 4 (Table 4) were used to introduce an amber codon at codon 180. Thus, CAG (gln) was changed to TAG (amber). Upon insertion of the cassette, this stop codon is removed and replaced with the wild-type residue glutamine. pTM04 does not complement DF502.

pTMCA was constructed by cassette mutagenesis (Richards, 1991). A double-stranded DNA cassette was synthesized from a single primer (primer 5, Table 4) that spanned the *Age*I and *Pm*I restriction sites. Primer 5 was synthesized with NNS (N = A, C, G, T; S = C, G) at the three C-terminal hinge codons 174–176. The primer was self-complementary at the 3'-end, and the 5'-end sequence mutated the *Pm*I site without changing the coding sense. Primer 5 (0.5 µg, 183 nmol) was denatured at 95 °C for 2 min, cooled to 16 °C over 3 h, and treated with Klenow fragment (10 units) in buffer A (40 µL). After polymerization, the cassette was restricted with *Age*I, purified by nondenaturing PAGE (12%, TBE

buffer), and phosphorylated with T4 polynucleotide kinase in buffer B. pTM04 (10 µg) was restricted with *Age*I (20 units), and *Pm*I (20 units) in buffer B, heat denatured for 10 min at 70 °C and dephosphorylated for 45 min at 37 °C. The cassette was ligated into the *Age*I-*Pm*I restricted vector with T4 DNA ligase (1,200 units) at 16 °C for 16 h in buffer B. The blunt end of the cassette was ligated to the blunt cut at the *Pm*I site of the vector. Thus, the library mutants no longer contained the *Pm*I site. The ligation mixture was desalting and electroporated (4–5 ms at 1.47 kV) into DH5α electrocompetent cells and plated onto LB/amp agar plates

GAC TGG AGT AAG GTG GTT CTT GCC TAT GAG CCG GTT TGG GCT ATC 510
D W S K V V L A W E P V W A I 170
<i>Age</i> I
GGA <b>ACC</b> GGT <b>AAA</b> ACT CGT ATC CCC CAA <b>TAG</b> GCT CAG AAG GTT CAT 555
G T G K T A T P Q amb A Q E V H 185
<i>Pm</i> I
GAG AAG CTG AGA GGC TGG CTC AAA AGC CAC <b>GTG</b> TCT GAT GCT GTT 600
E K L R G W L K S H V S D A V 200

**Fig. 9.** Gene and amino acid sequences of chicken triosephosphate isomerase pTM04 construct in hinge and loop regions. C-terminal hinge codons and residues are in boldface. Point mutations introduced for subcloning and reduction of wild-type background are underlined. Restriction sites utilized for cassette mutagenesis are italicized.

to generate the library. The DH5 $\alpha$  colonies on the plate were collected and combined. Two hundred thousand transformants were combined and plasmid purified to yield the pTMCA, the C-terminal hinge mutant library.

#### In vivo complementation and screening

To further reduce background wild-type contamination of the library, pTMCA (5  $\mu$ g) was digested with *PmI* (10 units) at 37 °C in buffer C for 12 h. The digested mixture that included wild-type and undesired deletion mutants was purified on a 1% LMP agarose gel, and the desired supercoiled bands were excised from the gel and purified using GeneClean III (BIO 101, Inc., Vista, California) following the manufacturer's protocol. Because the hinge mutants were not restricted (lacking the *PmI* site), only the mutants were supercoiled, circular plasmid DNA (Fig. 8B). The purified plasmid DNA was electroporated into DF502, recovered in a minimal medium without antibiotic (700  $\mu$ L) and plated onto minimal/amp/strep agar plates. Minimal plates were grown for 33 h at 37 °C. Colonies that were greater than 3 mm in diameter were selected and inoculated in 3 mL rich/amp/strep medium, and the cultures grown for 18 h at 37 °C. The cells were collected by centrifugation, frozen, and saved for plasmid DNA purification and specific activity assays.

A portion of the pTMCA library was transformed into DH5 $\alpha$ , individual colonies isolated, and their plasmid DNA sequenced in the region of the C-terminal hinge. Under these transformation conditions, selection for active mutants does not occur. Thus, the randomness of the library could be verified. The hinge region of 40 unselected mutants was sequenced. The proportion of each base occurring at each position is the same within one standard deviation as the percentage originally incorporated into the mutagenesis oligonucleotide. Consequently, the pTMCA library is not biased in base usage as a result of the library synthesis.

The electroporation efficiencies for DF502 and DH5 $\alpha$  were calculated by transforming each with a known quantity of pTM01. Typical efficiencies ranged from 10<sup>8</sup> transformants/ $\mu$ g to 10<sup>9</sup> transformants/ $\mu$ g. The concentration of pTMCA plasmid DNA transformed into DF502 was calculated by determining the number of transformants obtained upon electroporation of the same quantity of pTMCA into DH5 $\alpha$  and calculation from the previously determined efficiency of DH5 $\alpha$ .

#### Specific activity assay of triose phosphate isomerase mutants

All glassware and plasticware were either new or acid-washed to eliminate the possibility of contamination of triosephosphate isomerase from other sources. Cell paste (20 mg) of DF502(pTMCA) obtained from rich/amp/strep medium (2 mL) grown for 18 h was suspended in buffer D (1 mL), and lysed by two passages through a French press at 11,000 psi and 4 °C. Cell debris was removed by centrifugation at 15,000 g for 10 min. The amount of total protein in the supernatant was determined by Bradford assay (Bradford, 1976). The specific activity was determined by following the rate of GAP conversion at 30 °C. The assay mixture was comprised of 975  $\mu$ L buffer E, 5  $\mu$ L NADH (50 mg/mL), 5  $\mu$ L GOPDH (10 mg/mL), and 5  $\mu$ L of the crude lysate supernatant. The background rate of NADH oxidation was measured for the first 2 to 3 min. GAP (600  $\mu$ mol) was added to initiate the TIM-catalyzed reaction. The slope of the first 10% of the reaction was determined by linear regression and converted to mM min<sup>-1</sup> mg<sup>-1</sup> using

$\epsilon_{240} = 6,220 \text{ M}^{-1} \text{ cm}^{-1}$  (Horecker & Kornberg, 1948). Because triosephosphate isomerase is the major protein expressed, quantization of total protein corrects for variability in growth rates of transformed cells in liquid culture. Under these conditions, the specific activity of wild type is 55 mM min<sup>-1</sup> mg<sup>-1</sup>.

#### Electronic supplementary material

A full alignment of 62 TIM protein sequences is available (timseq).

#### Acknowledgments

This work was supported by a grant from the American Chemical Society–Petroleum Research Fund. The Center for Analysis and Synthesis of Macromolecules (CASM) at Stony Brook is supported by NIH Grant RR02427 and the Center for Biotechnology. We thank Ms. Mahbuba Yeasmin for technical assistance in studying amber codon-containing mutants, Prof. Daniel Raleigh for helpful discussions, and Prof. Hongshik Ahn for his invaluable assistance with the numerical simulations.

#### References

- Bairoch A, Boeckmann B. 1994. The SWISS-PROT protein sequence data bank: Current status. *Nucleic Acids Res* 22:3578–3580.
- Bradford MM. 1976. A rapid and sensitive method for the quantization of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* 72:248–254.
- Cserzö M, Simon I. 1989. Regularities in the primary structure of proteins. *Int J Peptide Protein Res* 34:184–195.
- de la Mare C, Coulson AFW, Knowles JR, Priddle JD, Offord RE. 1972. Active-site labelling of triose phosphate isomerase. *Biochem J* 129:321–331.
- Delboni LF, Mande SC, Rentier-Delrue F, Mainfroid V, Turley S, Vellieux FMD, Martial JA, Hol WGJ. 1995. Crystal structure of recombinant triosephosphate isomerase from *Bacillus stearothermophilus*. An analysis of potential thermostability factors in six isomerases with known three-dimensional structures points to the importance of hydrophobic interactions. *Protein Sci* 4:2594–2604.
- Derreumaux P, Schlick T. 1998. The loop opening/closing motion of the enzyme triosephosphate isomerase. *Biophys J* 74:72–81.
- Fetrow JS. 1995. Omega loops: Nonregular secondary structures significant in protein function and stability. *FASEB J* 9:708–717.
- Fraenkel D. 1986. Mutants in glucose metabolism. *Ann Rev Biochem* 55:317–337.
- Gerstein M, Lesk AM, Chothia C. 1994. Structural mechanisms for domain movements in proteins. *Biochemistry* 33:6739–6749.
- Gloss LM, Spencer DE, Kirsch JF. 1996. Cysteine-191 in aspartate aminotransferases appears to be conserved due to the lack of a neutral mutation pathway to the functional equivalent, alanine-191. *Proteins Struct Funct Genet* 24:195–208.
- Hermes JD, Blacklow SC, Knowles JR. 1990. Searching sequence space by definably random mutagenesis: Improving the catalytic potency of an enzyme. *Proc Natl Acad Sci USA* 87:696.
- Hermes JD, Parekh SM, Blacklow SC, Köster H, Knowles JR. 1989. A reliable method for random mutagenesis: The generation of mutant libraries using spiked oligodeoxyribonucleotide primers. *Gene* 84:143.
- Horecker BL, Kornberg A. 1948. The extinction coefficients of the reduced band of pyridine nucleotides. *J Biol Chem* 175:385–390.
- Joseph D, Petsko GA, Karplus M. 1990. Anatomy of a conformational change: Hinged “lid” motion of the triosephosphate isomerase loop. *Science* 249:1425–1428.
- Kempner ES. 1993. Movable lobes and flexible loops in proteins. Structural deformations that control biochemical activity. *FEBS Lett* 326:4–10.
- Knowles JR. 1991. Enzyme catalysis: Not different, just better. *Nature* 350:121–124.
- Knowles JR, Albery WJ. 1977. Perfection in enzyme catalysis: The energetics of triosephosphate isomerase. *Acc Chem Res* 10:105–111.
- Komives EA, Chang LC, Lolis E, Tilton RF, Petsko GA, Knowles JR. 1991. Electrophilic catalysis in triosephosphate isomerase: The role of histidine-95. *Biochemistry* 30:3011–3019.
- Kraulis PJ. 1991. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J Appl Crystal* 24:946–950.

- Leach SJ, Nemethy G, Scheraga HA. 1966. Computation of the sterically allowed conformations of peptides. *Biopolymers* 4:369–407.
- Leszczynski JF, Rose GD. 1986. Loops in globular proteins: A novel category of secondary structure. *Science* 234:849–855.
- Lolis E, Alber T, Davenport RC, Rose D, Hartman FC, Petsko GA. 1990. Structure of yeast triosephosphate isomerase at 1.9-Å resolution. *Biochemistry* 29:6609–6618.
- Lolis E, Petsko GA. 1990. Crystallographic analysis of the complex between triosephosphate isomerase and 2-phosphoglycolate at 2.5-Å resolution: Implications for catalysis. *Biochemistry* 29:6619–6625.
- Mande SC, Mainfray V, Kalk KH, Goraj K, Martial JA, Hol WG. 1994. Crystal structure of recombinant human triosephosphate isomerase at 2.8 Å resolution. Triosephosphate isomerase-related human genetic disorders and comparison with the trypanosomal enzyme. *Protein Sci* 3:810–821.
- Nemethy G, Leach SJ, Scheraga HA. 1966. The influence of amino acid side chains on the free energy of helix-coil transitions. *J Phys Chem* 70:998–1004.
- Noble MEM, Wierenga RK, Lambeir A-M, Opperdoes RR, Thunnissen A-MWH, Kalk KH, Groendijk H, Hol WGJ. 1991. The adaptability of the active site of trypanosomal triosephosphate isomerase as observed in the crystal structures of three different complexes. *Proteins Struct Funct Genet* 10:50–69.
- Noble MEM, Zeelen JP, Wierenga RK. 1993. Structures of the “open” and “closed” state of trypanosomal triosephosphate isomerase, as observed in a new crystal form: Implications for the reaction mechanism. *Proteins Struct Funct Genet* 16:311–326.
- Pompliano DL, Peyman A, Knowles JR. 1990. Stabilization of a reaction intermediate as a catalytic device: Definition of the functional role of the flexible loop in triosephosphate isomerase. *Biochemistry* 29:3186–3194.
- Raghava GPS, Sahni G. 1994. GMAP: A multi-purpose computer program to aid synthetic gene design, cassette mutagenesis and the introduction of potential restriction sites into DNA sequences. *BioTechniques* 16:1116–1123.
- Raines RT, Sutton EL, Straus DR, Gilbert W, Knowles JR. 1986. Reaction energetics of a mutant triosephosphate isomerase in which the active-site glutamate has been changed to aspartate. *Biochemistry* 25:7142–7154.
- Richards JH. 1991. Cassette mutagenesis. In: McPherson MJ, ed. *Directed mutagenesis. A practical approach*. New York: Oxford University Press. pp 199–215.
- Sambrook J, Fritsch EF, Maniatis T. 1989. *Molecular cloning: A laboratory manual*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.
- Sampson NS, Knowles JR. 1992. Segmental motion in catalysis: Investigation of a critical hydrogen bond for loop closure in the reaction of triosephosphate isomerase. *Biochemistry* 31:8488–8494.
- Straus D, Gilbert W. 1985. Chicken triosephosphate isomerase complements an *Escherichia coli* deficiency. *Proc Natl Acad Sci USA* 82:2014.
- Swinkels BW, Gibson WC, Osinga KA, Kramer R, Veeneman GH, van Boom JH, Borst P. 1986. Characterization of the gene for the microbody (glycosomal) triosephosphate isomerase of *Trypanosoma brucei*. *EMBO J* 5:1291–1298.
- Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673–4680.
- Velanker SS, Ray SS, Gokhale RSSS, Balaram H, Balaram P, Murthy MR. 1997. Triosephosphate isomerase from *Plasmodium falciparum*: The crystal structure provides insights into antimalarial drug design. *Structure* 5:751–761.
- Wierenga RK, Noble MEM, Vriend G, Nauche S, Hol WGJ. 1991. Refined 1.83 Å structure of trypanosomal triosephosphate isomerase crystallized in the presence of 1.4 M-ammonium sulphate. *J Mol Biol* 220:995–1015.
- Williams JC, McDermott AE. 1995. Dynamics of the flexible loop of triosephosphate isomerase: The loop motion is not ligand gated. *Biochemistry* 34:8309–8319.
- Zhang Z, Sugio S, Komives EA, Liu KD, Knowles JR, Petsko GA, Ringe D. 1994. Crystal structure of recombinant chicken triosephosphate isomerase-phosphoglycolohydroxamate complex at 1.8-Å resolution. *Biochemistry* 33:2830–2837.