ELSEVIER

Biophysical Chemistry 61 (1996) 131–141

Biophysical Chemistry

Thermodynamics of amino acid side-chain internal rotations

Andrew J. Doig

Department of Biochemistry and Applied Molecular Biology, UMIST, P.O. Box 88, Manchester M60 1QD, UK

Received 17 January 1996; accepted 11 March 1996

Abstract

The absolute Gibbs energy, enthalpy and entropy of each of the internal rotations found in protein side chains has been calculated. The calculation requires the moments of inertia of the side chains about each bond, the potential energy barrier and the symmetry number and gives the maximum possible thermodynamic consequences of restricting side chain motion when a protein folds. Hindering side chain internal rotations is unfavourable in terms of Gibbs energy and entropy; it is enthalpically favourable at 0 K. At room temperature, it is estimated that the adverse entropy of hindering buried side chain internal rotation is only 25% of the absolute entropy. The difference between absolute entropies in the folded and unfolded states gives the entropy change for folding. The estimated Gibbs energy change for restricting each residue correlates moderately well with the probability of that residue being found on the folded protein surface, rather than in the protein interior (where motion is restricted).

Keywords: Internal rotation; Side chain; Protein stability; Absolute entropy; Protein folding; Conformational entropy

1. Introduction

The forces which stabilise a folded protein relative to its unfolded state in aqueous solution remain poorly understood. These forces will include, at least, the hydrophobic effect, hydrogen bonding, van der Waals bonding in tightly packed cores and disulphide bond formation. Conformational strain in the folded state, and loss of conformational Gibbs energy upon folding, are effects which oppose folding. In this article, we investigate the thermodynamic properties of side chain internal rotations relative to a completely immobilised state and discuss possible implications for protein stability and structure. Consideration of side chain internal rotations, by an independent method, have recently been used by Creamer and Rose [1] to rationalise differences in α-helix propensities. Shakhnovich and Finkelstein [2] argued that the rate determining step in protein folding is finding the correct side chain rotamers.

Fig. 1 shows the internal rotations present in the side chains of the 20 naturally occurring amino acid side chains in proteins. Rotation is assumed to be possible about each of these bonds in the unfolded protein and may be restricted in the folded state by close packing and specific side chain-side chain interactions, such as hydrogen bonding and salt-bridges. Side chains that remain exposed to solvent on the protein surface may have some mobility restricted by secondary structure formation, but will, in general, lose less conformational entropy than buried side chains. The Gibbs energies, enthalpies

PII S0301-4622(96)02180-1

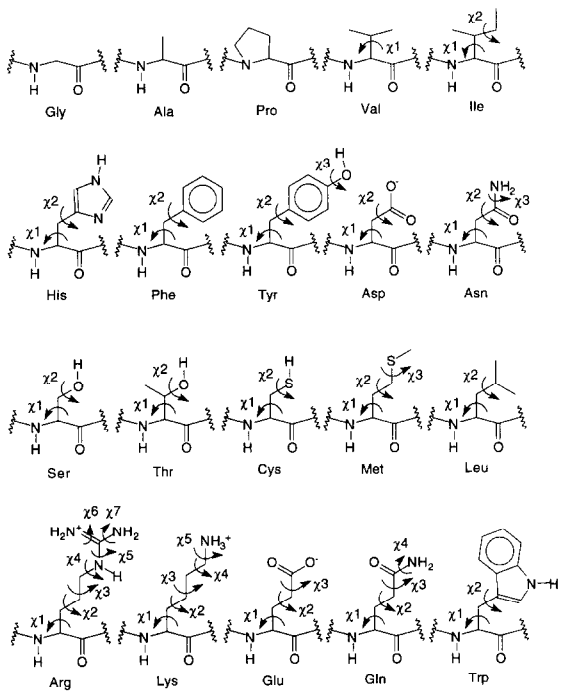

Fig. 1. Internal rotations present in the twenty naturally occurring side chains.

and entropies of each of the rotors shown in Fig. 1 can be calculated from theory.

2. Materials and methods

The theory used here was derived in a series of papers by Pitzer and co-workers [3–6]. A useful summary of the theory is given in Ref. [7]. The partition function (Q_f) for a free rotation (i.e., an internal rotation with no energy barrier, where the potential energy is invariant with dihedral angle) is given by Eq. (1).

$$Q_f = 2.7935 \frac{\sqrt{(10^{38} g^{-1} cm^{-2} K^{-1}) I_r T}}{n} \quad (1)$$

where I_r is the reduced moment of inertia about the bond axis in $g\,cm^2$, T is the absolute temperature (here 298 K) and n is the symmetry number of the rotation (the number of identical structures per 360° rotation). I_r is calculated by measuring the moment of inertia of the side chain about the bond axis for the atoms attached (directly or indirectly) to the terminal end of the side chain only (Fig. 2). The assumption of ignoring the remainder of the protein on the backbone side of the bond axis can be justified by considering the symmetric, coaxial top equation $I_r = I_1 I_2/(I_1 + I_2)$ (where I_1 and I_2 are the moments of inertia of the two groups on either end of the bond), which is approximate for the asymmetric side chains and is a simplification of the general case. If the group on one end of the bond is very large (e.g., a polypeptide backbone) and the group on the other end is comparatively small (e.g., part of one side chain) then $I_1 \gg I_2$ and $I_r \approx I_1 I_2/(I_1 + I_2) \approx I_2$. The more general, exact equation for the reduced moment of inertia [4] can be simplified similarly. It is therefore not necessary to know the moment of inertia of the very large polypeptide backbone; only the moment of inertia of the smaller group is needed. This is fortunate as estimating the moment of inertia of almost all of an unfolded protein about one axis is a formidable problem, because of the vast number of conformations that an unfolded protein may adopt. Ignoring the larger group (I_1) is intuitively reasonable as the rate of an internal rotation will depend more on the size of the smaller group attached to one end of the bond axis rather than the larger group. Thus, the rate at which a dog's tail wags depends on the size of the tail, not on the size of the dog.

The Gibbs energy of an internal rotation with no potential barrier is $-RT \ln Q_f$. This is now corrected to take into account the potential energy barrier, a procedure that will always make the Gibbs energy smaller in magnitude. The correction is a

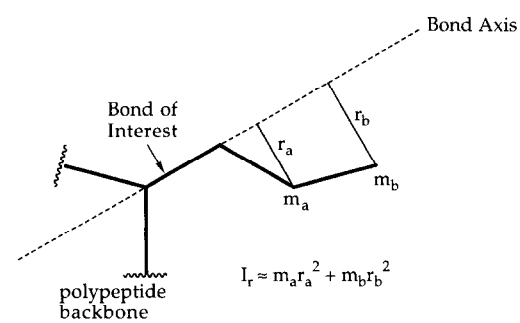

Fig. 2. Calculation of reduced moments of inertia, I_r, for a two atom side chain.

Table 1
Moments of inertia, potential barriers, Gibbs energies, enthalpies and entropies of side chain internal rotations at 298 K

Residue and bond no.	M. Of I. (10^{-40}g · cm^2)	Barrier (kJ · mol^{-1})	Enthalpy (kJ · mol^{-1})	Entropy (J · K^{-1}mol^{-1})	Free Energy (kJ · mol^{-1})	Total Free Energy (kJ · mol^{-1})
Ser χ2	0.84	4.8	1.1	15	−3.0	−8.1
Ser χ1	32	16.4	2.1	24	−5.1	
Cys χ2	1.8	6.8	1.4	18	−3.9	−10.0
Cys χ1	98	18.9	2.2	28	−6.1	
Thr χ2	0.84	4.7	1.1	15	−3.0	−8.6
Thr χ1	69	20.8	2.1	26	−5.6	
Arg χ7	1.5	50.0	0.54	2.5	−0.2	−32.8
Arg χ6	1.5	49.9	0.54	2.5	−0.2	
Arg χ5	110	49.0	2.0	15	−2.5	
Arg χ4	120	17.9	2.3	29	−6.4	
Arg χ3	450	17.6	2.5	35	−7.9	
Arg χ2	250	25.0	2.3	30	−6.8	
Arg χ1	1010	18.0	2.5	38	−8.8	
Phe χ2	90	13.4	2.5	24	−4.6	−13.1
Phe χ1	720	16.5	2.5	37	−8.5	
Tyr χ3	0.79	15.0	0.6	12	−2.9	−16.5
Tyr χ2	92	13.4	2.5	24	−4.6	
Tyr χ1	1200	16.5	2.5	39	−9.0	
Met χ3	54	14.5	2.2	27	−5.8	−19.3
Met χ2	120	19.5	2.3	28	−6.2	
Met χ1	270	18.0	2.4	33	−7.3	
Val χ1	72	22.5	2.1	26	−5.6	−5.6
Trp χ2	440	12.0	2.6	37	−8.4	−17.8
Trp χ1	1520	16.8	2.5	40	−9.4	
His χ2	55	5.3	2.4	32	−7.1	−15.2
His χ1	520	17.3	2.5	35	−8.1	
Ile χ2	37	21.1	2.6	23	−5.0	−10.9
Ile χ1	110	24.4	2.2	27	−5.9	
Leu χ2	73	22.5	2.1	26	−5.6	−12.0
Leu χ1	160	22.9	2.3	29	−6.4	
Asn χ3	1.5	66.2	0.36	1.5	−0.1	−13.3
Asn χ2	44	8.3	2.4	29	−6.3	
Asn χ1	200	19.1	2.3	30	−6.9	
Gln χ4	1.5	66.2	0.36	1.5	−0.1	−20.4
Gln χ3	44	5.6	2.4	31	−6.8	
Gln χ2	200	19.8	2.3	31	−6.8	
Gln χ1	170	18.3	2.3	30	−6.7	
Asp χ2	40	5.5	2.4	25	−4.9	−11.7
Asp χ1	180	18.5	2.3	31	−6.8	
Glu χ3	40	2.5	2.0	26	−5.8	−19.3
Glu χ2	190	19.8	2.3	30	−6.8	
Glu χ1	170	18.2	2.3	30	−6.7	
Lys χ5	2.7	11.9	1.3	7	−0.8	−25.7
Lys χ4	39	16.7	2.1	25	−5.3	
Lys χ3	75	19.8	2.2	27	−5.8	
Lys χ2	210	26.0	2.3	30	−6.3	
Lys χ1	320	18.0	2.4	33	−7.5	
Mean	204	20.3	2.0	26	−5.6	−15.4

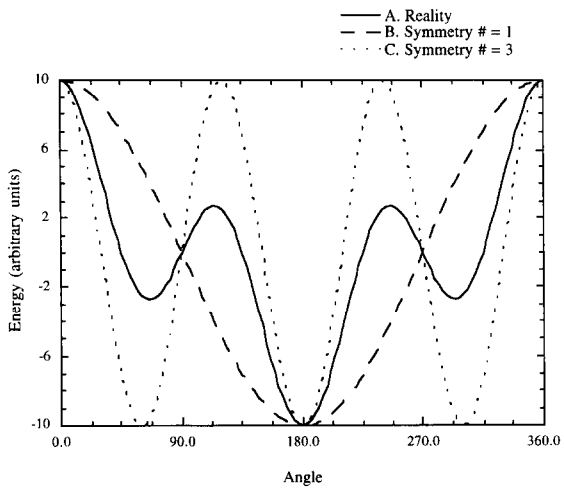

Fig. 3. Curves of potential energy versus dihedral bond angle. Curve A shows an hypothetical potential that has three minima, though at different energies. Curves B and C show how the theory can be used to approximate this potential. Curve B has a symmetry number of 1 and curve C has a symmetry number of 3.

function of the barrier height and partition function [7]. This correction factor assumes that the potential energy curve is a cosine. The effects of deviations from this ideal are discussed below. The entropy of an internal rotation is given by Eq. (2).

$$S = R(0.5 + \ln Q_f) - (S_f - S) \quad (2)$$

S = free rotation entropy-barrier correction, where $R(0.5 + \ln Q_f)$ is the entropy of the free rotation and $(S_f - S)$ is the correction to take account of the energy barrier [7]. The enthalpy is available from Pitzer's tables as a function of the partition function and the barrier. For a free internal rotation with no barrier, the results simplify to $H = 0.5RT$ and $S = R(0.5 + \ln Q_f)$.

As mentioned above, the theory assumes that the potential energy curve is a cosine. The theory can deal with deviations from an exact cosine potential in two ways. Fig. 3 shows an observed potential energy curve (A) with an approximate, but not exact three-fold symmetry. The theory can approximate this curve in one of two ways. Either a cosine curve is used that has a symmetry number of 1 (curve B in Fig. 3) or a symmetry number of 3 (curve C in Fig. 3). Both these alternatives have the same potential energy barrier (difference between greatest peak and lowest trough) as the observed curve A. The thermodynamic functions for the curve B approximation is calculated as normal with a symmetry number of 1. For the curve C approximation, the thermodynamics of the internal rotation are calculated with a symmetry number of 3. If the true function is not exactly symmetric, R ln 3 is then added to the entropy and $-RT$ ln 3 to the Gibbs energy; the enthalpy remains unchanged. Two sets of results were found using both methods. Method 1 used a symmetry number of 1 (i.e., curve B in Fig. 3) and Method 2 used a symmetry number of 3 (curve C in Fig. 3). The results found using Method 2 are given in Table 1.

Preliminary work on CH_3-CONH-CHR-CONH-CH_3 peptides (where R is the side chain of the amino acid under investigation), using MacroModel [8] and the AMBER force field [9] suggests that a cosine function with a symmetry number of 3 is more accurate. For example, the 3 conformations of lowest energy about the internal rotation differ by generally only about 10% of the barrier (results not shown). However, this is a somewhat circular argument as the force field uses a cosine function with three-fold symmetry for dihedral energy terms. Deviations from an exact cosine originate in longer range interactions such as electrostatics and van der Waals clashes. A more complex theory is required to take into account the exact shape of the potential function, though obtaining such a function precisely is not a trivial problem. In the remainder of this article the results quoted will be those obtained using Method 2, as this is probably more accurate. In any case, the results found with the two methods are very similar, especially for the entropies. The results found using Method 2 are given in Table 1. The Method 1 results are not shown.

A second problem in these calculations is in dealing with the potential effects of solvent water binding to polar groups. If this occurs throughout an internal rotation, the moment of inertia of that internal rotation will be increased and hence also the Gibbs energy, enthalpy and entropy. Non-polar groups will presumably be unaffected. The likelihood of water remaining bound during an internal rotation can be assessed from the lifetimes of water hydrating unfolded protein compared to the time required for internal rotations. Studies on unfolded performic acid-oxidised ribonuclease A by ^{13}C NMR have found that the correlation time for side chains is

in the range 0.2 to 1.6 ns [10]. The correlation time for rotations closer to the backbone, such as Lys $\chi 1$ compared with Lys $\chi 4$, is larger, as expected. The correlation time for backbone internal rotations is about 2 ns. The residence times for water on the protein surface, which is presumably similar to that in the unfolded state, has been measured by NMR [11]. It was found that all surface waters had residence times in the subnanosecond range. The lifetime of a water molecule solvating a polar group is therefore similar to the time required for an internal rotation. The similar timescales for internal rotations and water lifetimes attached to polar groups suggests that an internal rotation may take place each time the water-side chain hydrogen bond weakens by the water diffusing away. Since there does not appear to be any evidence for long lifetimes of water bound to surface side chains we will assume here that bound waters have no significant effect on moments of inertia of side chain internal rotations, though this problem is not yet resolved completely.

3. Results

The potential energy barriers to the rotations were calculated using MacroModel [8], the AMBER force field [9] and a dielectric constant of 80, a value appropriate for rotations in water, at 298 K. Each of the amino acids was placed in a CH_3-CONH-CHR-CONH-CH_3 molecule (where R is the side chain of the amino acid under investigation). An uncharged molecule was used to avoid strong ionic interactions between the side chain and the backbone that would be present in the unfolded state (except close to the N- and C-termini). The structure was energy minimised, starting from β-sheet backbone dihedral angles and extended side chains. Typically $\chi_1 = 60°$ and χ_2, χ_3 and $\chi_4 = 180°$. This initial conformation was allowed to relax into a nearby local minimum by energy minimisation and the energy of the relaxed conformation calculated. The dihedral angle of the bond being considered was then constrained into a high energy, eclipsed conformation while the remainder of the molecule was allowed to relax into a nearby low energy conformation to remove too-close van der Waals contacts, etc. The energy of the molecule was calculated, after energy minimisation and then the removal of the constraint. The potential energy barrier to the internal rotation is the difference between this energy and the energy of the energy minimised molecule. All the possible eclipsed conformers were investigated to ensure that the highest energy conformer had been found. The results obtained are given in Table 1. Potential barriers were also calculated in an identical manner except using a dielectric constant of 1, the dielectric of a vacuum. It was found that while the absolute energies varied greatly from those calculated with a dielectric of 80, the energy barriers, which are the differences between two absolute energies, agreed to within 10% in the few cases considered (results not shown). This is because the dielectric is not part of the dihedral angle energy term in the AMBER force field which dominates these energy changes. Individual water molecules bound to polar groups were not considered when estimating the potential barriers. A more accurate method for determining the internal energy as a function of dihedral angle has been described [12].

The moments of inertia of the terminal ends of the side chains were calculated using the energy minimised structure described above and a FORTRAN program which used a MacroModel data file. Moments of inertia for 18 internal rotations have been published by Lee et al. [13]. Our results are virtually identical.

The results obtained are shown in Table 1. Clearly, the side chains will not always be in their lowest energy conformations during an internal rotation which introduces an error into the estimates of the moments of inertia. The symmetry numbers of the internal rotations are 1 apart from the following: Arg $\chi 7$ (2), Arg $\chi 6$ (2), Arg $\chi 5$ (2), Phe $\chi 2$ (2), Tyr $\chi 2$ (2), Asn $\chi 3$ (2), Gln $\chi 4$ (2), Asp $\chi 2$ (2), Glu $\chi 3$ (2) and Lys $\chi 5$ (3).

It is notable that the internal rotations that are restricted by conjugation (such as those within the guanidinium group) have much smaller Gibbs energies. For example, Asn $\chi 3$ has a Gibbs energy of only $-0.1\ kJ\ mol^{-1}$ because of its small moment of inertia resulting from the rotation of two hydrogens about the C–N bond axis, a high barrier because of conjugation and a symmetry number of 2. Similarly, Arg $\chi 5$ has a Gibbs energy of only $-2.5\ kJ\ mol^{-1}$, despite its sizeable moment of inertia of $49 \times$

10^{-40} g cm^2, because of its high barrier and symmetry.

Crankshaft motion can occur when a small central portion of a molecule rotates about two or more single bonds, thus avoiding large motions at the ends of the chain. Its expected prevalence within the protein backbone means that the methods here cannot be used for backbone internal rotations. It is likely that some co-operative crankshaft rotation will take place within some of the larger side chains. If this is significant, the Gibbs energies of the internal rotations will be less certain (perhaps smaller), particularly for the internal rotations close to the backbone of the longer side chains. It is also possible that the values given here are overestimates for aqueous solution, as a fairly viscous solvent, such as water, is likely to decrease the rate of internal rotation and make it easier for such motions to be restricted. Such effects cannot be quantified using the present theory. Hydrogen bonding of polar side chains back to the main chain or interactions of side chains with each other may also lower rotational entropies by hindering motion in the unfolded state.

Although there are a number of approximations in these calculations (discussed above), the thermodynamic quantities obtained for rotations in the unfolded state (Table 1) are often relatively insensitive to the assumptions made and to the different moments of inertia or potential barriers. For example, Tyr $\chi 1$ and Ser $\chi 1$ have almost identical potential barriers while the moment of inertia of Tyr $\chi 1$ is 38 times larger than that of Ser $\chi 1$. Despite this large range, the Gibbs energy of Tyr $\chi 1$ is only 1.8 times larger than that of Ser $\chi 1$. This is primarily because the Gibbs energy depends only on the logarithm of the reduced moment of inertia. Similarly, Met $\chi 3$ and His $\chi 2$ have very similar moments of inertia but the barrier of Met $\chi 3$ is 2.7 times large than that of His $\chi 2$. This difference, however, results in the Gibbs energy of His $\chi 2$ being only 1.2 times larger than that of Met $\chi 3$.

As the pK_a of His is 7, both protonated and unprotonated His are commonly present in proteins. The extra proton will give slightly different moments of inertia to the two forms and hence slightly different results for the internal rotations. These were calculated for protonated and unprotonated His but no significant differences were found to within 0.1 kJ mol^{-1} for G and H and 1 J K^{-1} mol^{-1} for S. The results given in Table 1 for His are therefore appropriate for both states of His. The protonated and unprotonated states have not been compared for any other side chains.

In a preliminary version of this article we calculated the thermodynamic properties of the internal rotations assuming that bound waters were present [14]. Comparison of the results presented here with those published previously illustrate the effects of bound waters. In the previous work the following numbers of water molecules were assumed to rotate with the following functional groups: CO_2^- (5), NH_3^+ (3), $CONH_2$ (4), uncharged imidazole (2), charged imidazole (4), guanidine (5), SH (1), OH (1) and NH (1). The effect of these bound waters on the Gibbs energies of the internal rotations varied from site to site. Where the moment of inertia without water is very small, such as Ser $\chi 2$, the addition of a single water to the hydroxyl group increased the moment of inertia 180-fold from 0.84×10^{40} g cm^2 to 150×10^{40} g cm^2 and hence the Gibbs energy nearly 3-fold from -3.0 kJ mol^{-1} to -8.4 kJ mol^{-1}. The addition of water had a much smaller effect when the moment of inertia of the group without bound water was larger. For example the addition of 5 waters increased the moment of inertia of Arg $\chi 1$ from 1010×10^{40} g cm^2 to 3700×10^{40} g cm^2 and the Gibbs energy from -8.8 kJ mol^{-1} to -10.3 kJ mol^{-1}. The uncertainty in the numbers and locations of these bound waters makes the results of these calculations on polar side chains less reliable than for non-polar side chains. As non-polar side chains will not have any bound waters, if bound water is important, only the Gibbs energies of the polar side chains will be increased. Bound waters increased the total Gibbs energies of the polar side chains by about 50% on average.

The enthalpies in Table 1 are positive thus showing that restriction of side chain motion is enthalpically favourable. This is because the hindered rotation will be on average closer to the bottom of the potential energy well. The restricting of internal rotations is unfavourable in terms of Gibbs energy and entropy as a smaller range of dihedral angles can now be sampled.

Lee et al. [13] calculated absolute entropies of residues X in an $(Ala)_4$-X-$(Ala)_4$ α-helical peptide,

Table 2
Entropy change for folding a side chain from the unfolded state to a helix surface

Side Chain	S (unfolded) [a] ($J \cdot K^{-1} mol^{-1}$)	S (helix surface) [b] ($J \cdot K^{-1} mol^{-1}$)	ΔS ($J \cdot K^{-1} mol^{-1}$)
Val	26	16	−10
Leu	55	44	−11
Ile	50	39	−11
Phe	61	53	−8
Trp	77	68	−9
His	67	58	−9
Asp	56	48	−8
Ser	39	35	−4
Thr	41	36	−5

[a] Data from Table 1. Sum of $\chi 1$ and $\chi 2$.
[b] Data from Ref. [13]. Cys is not consider as Lee et al. considered $\chi 1$ only while we consider $\chi 1$ and $\chi 2$ for this side chain.

where X is a short residue. The difference between their entropies and the results here gives the entropy change for transferring a side chain from the unfolded state to the surface of an α-helix. For example, the absolute entropy of Val in the unfolded state is $26 J K^{-1} mol^{-1}$ (Table 1), the absolute entropy of Val on the surface of an α-helix is $16 J K^{-1} mol^{-1}$ [13], giving the entropy change for folding the Val side chain as $-10 J K^{-1} mol^{-1}$. The complete set of these calculations is given in Table 2. Lee et al. stated that "The value of the configurational entropy is difficult to calculate in the unfolded state..." and hence only made calculations for side chains in isolated folded helices. The work presented here therefore nicely complements their work as it enables the entropy change for folding to be calculated.

4. Discussion and conclusion

The results listed in Table 1 should be reasonable estimates of the Gibbs energies, enthalpies and entropies of side chain internal rotations in the unfolded state relative to a stationary rotor at the bottom of its potential energy well (i.e., 0 K). The estimated entropies of side chain internal rotations are therefore Third Law (absolute) entropies. However, within a protein core, or any complex involving peptides, there will be some residual motion about the bonds, such as in the form of a torsional vibration, which will retain some of the Gibbs energy, entropy and enthalpy of the internal rotation in the unfolded or unbound state. This has been experimentally verified by the NMR observation that some groups will continue to rotate at a significant rate (e.g., the aromatic rings of tyrosine or phenylalanine about their C_1–C_4 axes) within the folded protein on occasion, leading to averaged proton signals [15]. It is therefore important to stress that the results listed in Table 1 can give upper limits only of the thermodynamic changes that occur in protein folding and other situations where internal rotations are hindered. All the calculated Gibbs energy can be lost only if an infinitely high energy barrier is imposed or the group is cooled to absolute zero.

It is possible to divide the absolute configurational entropy, calculated here, into two parts. Firstly, there will be a change in the number of rotamers populated (conformational entropy); secondly, the width of an allowed potential energy well might narrow (vibrational entropy). Karplus et al. [16] used molecular dynamics simulations to estimate vibrational and conformational entropy during folding. They suggested that although the magnitude of absolute vibrational entropy is nearly an order of magnitude larger than conformational entropy, the most important change on folding is in conformational entropy.

A number of groups have recently made estimates of side chain conformational entropies (reviewed in Ref. [17]). A consensus was reached that the entropy change on folding is approximately $-14 J K^{-1} mol^{-1}$ per residue in protein folding and $\approx -7 J K^{-1} mol^{-1}$ per buried rotamer [17]. These figures can be compared to the mean absolute entropy of $26 J K^{-1} mol^{-1}$ per rotor (Table 1). This shows that only $\approx 25\%$ of the absolute entropy of the internal rotation is lost in protein folding. A larger fraction may be lost where the constraints are greater such as during covalent bond or transition state formation [18]. Polar interactions within folded proteins may provide such a more restricted environment. If the barrier within a folded protein can be measured or calculated, the methods used here may be suitable for estimating the thermodynamics of these residual motions, though motion in this case may be more complex than in the unfolded state. If the torsional vibration in the folded protein is not about the minimum energy conformation, this will

contribute to enthalpically unfavourable strain energy.

Entropy changes for folding a side chain onto the surface of an α-helix are given in Table 2. It is seen that the observed entropy change is, on average, 16% of the absolute entropy. A larger fraction of the absolute entropy will presumably be lost for folding into a more restricted hydrophobic core. These results are somwhat larger than those of Creamer and Rose [1] who calculated this quantity by an independent method.

The negative enthalpy change that results from the complete restriction of internal rotations is an effect on the enthalpy of protein folding that has not been previously noted, to our knowledge. Table 1 shows that this effect is at most $\approx -2\,\mathrm{kJ\,mol^{-1}}$ per internal rotation at 0 K. At room temperature, it will be offset (perhaps even fully) by the enthalpy of torsional vibrations in the folded protein. Any residual enthalpy change originating in this way will also be offset by strain. Each internal rotation has a heat capacity (C_V) of approximately $8\,\mathrm{J\,K^{-1}\,mol^{-1}}$ [7]. A change in heat capacity from restriction of motion is therefore possible, though it is probably small, as heat capacity changes in protein folding appear to be dominated by changes in hydration [19–23].

It is noteworthy that hydrophobicity and the effect of the restriction of internal rotations act in the opposite sense for non-polar side chains. As a non-polar side chain increases in length, the hydrophobic effect becomes more favourable. However, a longer side chain will generally have a larger number of internal rotations, which will discourage the burial of such a group within a solid-like protein core. This will be offset by improved van der Waals bonding in a denser protein core [24] and opposed by the introduction of strain. Note that the sum of the Gibbs energies of internal rotations for side chains will not be simply proportional to the solvent accessible surface area of the side chain since internal rotations depend upon shape as well as size. The contribution a non-polar residue makes to the Gibbs energy of protein folding is therefore not expected to be exactly proportional to surface area, though this is widely assumed. The thermodynamic effects of close packing and restriction of side chain internal rotations have been considered by Nicholls et al. [25] who used hydrocarbon melting as a model for hydrophobic core formation from a liquid hydrocarbon state. The energetics of aqueous solution to liquid hydrocarbon transfer (the hydrophobic effect) can be assessed by solvent transfer experiments. They concluded that close packing and side chain restriction come close to cancelling one another out.

The relative importance of internal rotations and the hydrophobic effect in protein structure can be investigated by examining these quantities for each residue and seeing how well they correlate with protein structure. All other things being equal, it is expected that side chains with a higher number of internal rotations will be less likely to be buried as the restriction of internal rotations is unfavourable. An 'equilibrium constant' for partitioning between a protein surface and interior can be obtained from the fraction of residues buried. The logarithm of this equilibrium constant is proportional to a Gibbs energy for transfer between the protein surface and interior (ΔG_{tr}). These transfer Gibbs energies have been calculated by Radzicka and Wolfenden [26], based upon a survey of protein crystal structures by Chothia [27], who defined a buried residue as one that has $< 5\%$ of the surface area of the residue in a Gly-X-Gly tripeptide. If ΔG_{tr} is plotted against the cost in Gibbs energy of completely freezing out the rotors in a side chain (ΔG_{rot}), the correlation is reasonable ($R = 0.76$, Fig. 4), though generally poorer than in plots of ΔG_{tr} versus hydrophobicity [26]. Pickett and Sternberg [28] similarly concluded that an improved correlation between ΔG_{tr} and hydrophobicity can be reached if side chain entropy is added to the hydrophobicity scale. Alternative definitions of surface exposed and buried residues will give somewhat different plots. We previously made this plot using internal rotation Gibbs energies obtained by including water molecules bound to polar side chains [14]. A better correlation was obtained then because the Gibbs energies of the internal rotations in polar side chains were larger. The difference between polar and non-polar side chains was thus exaggerated.

Non-polar residues are more likely to be buried in folded proteins than to be found on the protein surface; polar residues show the opposite trend. This observation has traditionally been explained by the fact that it is more favourable to transfer a non-polar residue to a protein core than a polar residue, as a

result of the hydrophobic effect [29]. However, while this is clearly a real and important effect, the evidence of Fig. 4 suggests the partitioning of residues between the protein core and the protein surface may also be affected by the energetic cost of restricting the internal rotations in a side chain, though to a lesser extent. The folding of polar residues is discouraged as a result of polar groups generally being located on the ends of long, unbranched flexible chains with many internal rotations. Non-polar residues are generally easier to bury as they are short and branched. Branching decreases the number of internal rotations in a side chain and therefore makes burial easier. The overall contribution of restricting side chain motion to protein stability may be relatively small as (i) residues with many internal rotations tend to stay on the protein surface and (ii) only 25% of the absolute entropy of the internal rotation may be lost upon folding (see earlier). Nevertheless, a mean value of approximately $14\,J\,K^{-1}\,mol^{-1}$ per residue in protein folding is still important.

A better correlation is generally found between the degree of side chain burial in proteins and hydrophobicity [26], than between burial and internal

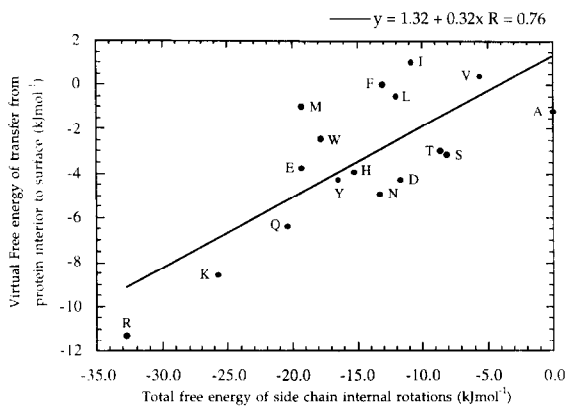

Fig. 4. Correlation between Gibbs energy of side chain internal rotations (ΔG_{rot}) and virtual Gibbs energy of transfer from protein surface to interior (ΔG_{tr}). In this graph the residues Cys, Pro and Gly are not included. Cys is not included as it is found commonly in disulphide bonds and the cost of freezing a disulphide unit is unknown and cannot be calculated by the method used here as the side chains form part of a loop. Gly and Pro are not included as the internal rotations in their backbones differ greatly from the other residues [33]. Gly is more flexible than other residues and Pro has only a single backbone rotor caused by the 5-membered ring it possesses.

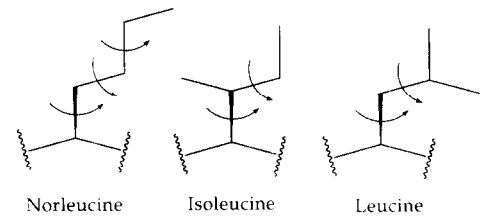

Fig. 5. Internal rotations in norleucine, isoleucine and leucine.

rotations. This supports the belief that internal rotations are less important to protein structure and stability than hydrophobicity. Gibbs energies of side chain internal rotations correlates moderately well with hydrophobicity. ΔG_{rot} was plotted against residue transfer data from cyclohexane [26], vapour [30], and 1-octanol [31] to aqueous solution (not shown). The least squares correlation coefficients found were 0.61, 0.74 and 0.37, respectively. Thus, more polar side chains generally have more internal rotations.

Consideration of internal rotations may give some insight into why these particular side chains were selected by evolution. Residues such as Val, Ile and Leu form protein cores readily not only because of their hydrophobic nature but also because their branched side chains decrease the number of internal rotations that need to be restricted in a protein core. For example, the amino acid norleucine (Fig. 5), which will have a similar hydrophobicity to leucine and isoleucine, is not found in proteins. Isoleucine and leucine have the advantage over norleucine of having one fewer internal rotation. The burial of the naturally occurring residues is therefore energetically favoured. Similarly, charged residues, such as arginine and lysine, have their polar groups on the ends of long, unbranched chains. Their burial in protein cores is discouraged not only by their polar nature but also by the large number of internal rotations which will be restricted. If groups such as Arg and Lys are buried they must form strong interactions to overcome the large loss of entropy from restriction of their motion.

The energetics of protein folding are often assessed by solvent transfer experiments where compounds which model protein side chains are transferred from aqueous to non-polar solution. However, a non-polar solution is an inappropriate model for a

protein core as it is not close-packed. The effects of improved van der Waals bonding and restriction of internal rotations in a close-packed core cannot be observed in solution. Side chains such as norleucine and isoleucine may show very similar energetics of solvent transfer and van der Waals bonding, as they have similar surface areas, but the cost of restricting their internal rotations will differ.

It has been shown that the interior of membrane bound proteins is comparable in hydrophobicity with that of water soluble proteins, rather than being 'inside-out' with polar groups buried and non-polar groups on the surface, interacting with the non-polar solvent [32]. This observation suggests that determining whether a particular residue is more likely to be buried in a membrane protein is dominated by the cost of restricting the internal rotations in the side chain, rather than the hydrophobicity of the residue, since for unfolding in a membrane there is a negligible hydrophobic effect. The authors suggested that the interior of membrane bound proteins are non-polar as "van der Waals forces in close-packed interiors of all proteins are crucial for protein stability". This is very likely true; however, there is no obvious reason why close-packed interiors cannot be formed using some polar atoms and hence why polar groups do not form part of protein cores. The most important determinant of which residues tend to form the interior of membrane-soluble proteins may be the cost of restricting the internal rotations in the residue side chain, an unfavourable Gibbs energy which will be relevant in any solvent.

In conclusion, Pitzer's theory was applied to amino acid side chain internal rotations to calculate their thermodynamic properties relative to complete immobilisation at 0 K. The Gibbs energy of an internal rotation increases with the moment of inertia about the bond axis and decreases with the height of the barrier to the rotation, increasing symmetry and with the number of water molecules bound to the side chain during the rotation. Restricting an internal rotation is unfavourable in terms of Gibbs energy and entropy but enthalpically favourable.

Acknowledgements

I thank Dudley Williams, Mark Searle, Mike Sternberg, Robert Baldwin, Doug Laurents, Steve Mayo, Trevor Creamer and Ben Stapley for helpful discussions. Mark Gardner is thanked for computer programming and Ute Gerhard for assistance with calculations. I also thank the BBSRC and the Nuffield Foundation for financial support.

References

[1] T.P. Creamer and G.D. Rose, Proc. Nat Acad. Sci. USA, 89 (1992) 5937–5941.
[2] E.I Shakhnovich and A.V. Finkelstein, Biopolymers, 28 (1989) 1667–1680.
[3] K.S. Pitzer and W.D. Gwinn, J. Chem. Phys., 10 (1942) 428–440.
[4] K.S. Pitzer, J. Chem. Phys., 14 (1946) 239–243.
[5] J.E. Kilpatrick and K.S. Pitzer, J. Chem. Phys., 17 (1949) 1064–1075.
[6] K.S. Pitzer, J. Chem. Phys., 5 (1937) 469–472.
[7] K.S. Pitzer, Quantum Chemistry. Constable, London, 1953, pp. 492–500.
[8] F. Mohamadi, N.G.J. Richards, W.C. Guida, R. Liskamp, M. Lipton, C. Caufield, G. Chang, T. Hendrickson and W.C. Still, J. Comp. Chem., 11 (1990) 440–467.
[9] S.J. Weiner, P.A. Kollman, D. Case, U.C. Singh, G. Alagona, S. Profeta and P. Weiner, J. Am. Chem. Soc., 106 (1984) 765–784.
[10] V. Glushko, P.J. Lawson and F.R.N. Gurd, J. Biol. Chem., 247 (1972) 3176–3185.
[11] G. Otting, E. Liepinsh and K. Wüthrich, Science, 254 (1991) 974–980.
[12] R.C. Wade and J.A. McCammon, J. Mol. Biol., 225 (1992) 679–696.
[13] K.H. Lee, D. Xie, E. Freire, and L.M. Amzel, L.M. Proteins: Structure, Function, and Genetics, 20 (1994) 68–84.
[14] A.J. Doig,, M.S. Searle,, M. Gardner and D.H. Williams, Techniques in Protein Chemistry IV, Academic Press, San Diego, CA, 1993, pp. 557–566.
[15] T.E. Creighton, Proteins, 1st edn., Freeman, New York, 1984, pp. 284–286.
[16] M. Karplus, T. Ichiye and B.M. Pettitt, Biophys. J. 52 (1987) 1083–1085.
[17] A.J. Doig and M.J.E. Sternberg, Prot. Sci., 4 (1995) 2247–2251.
[18] M.I. Page and W.P. Jencks, Proc. Nat. Acad. Sci. USA, 68 (1971) 1678–1683.
[19] R.L. Baldwin, Proc. Nat. Acad. Sci. USA, 82 (1986) 8069–8072.
[20] J.T. Edsall, J. Am. Chem. Soc., 57 (1935) 1506–1507.
[21] P.L. Privalov, Adv. Protein Chem., 33 (1979) 167–241.
[22] P.L. Privalov, and S.J. Gill, Adv. Protein Chem., 39 (1988) 191–234.
[23] D.N. Woolfson, A. Cooper, M.M. Harding, D.H. Williams and P.A. Evans, J. Mol. Biol., 229 (1993) 502–511.
[24] J. Bello, Int J. Pept. Protein Res., 12 (1978) 38–41.

[25] A. Nicholls, K.A. Sharp and B. Honig, Proteins: Struct., Funct., Genet., 11 (1991) 281–296.
[26] A. Radzicka and R. Wolfenden, Biochemistry, 27 (1988) 1664–1670.
[27] C. Chothia, J. Mol. Biol., 105 (1976) 1–12.
[28] S.D. Pickett and M.J.E. Sternberg, J. Mol. Biol., 231 (1993) 825–839.
[29] W. Kauzmann, Adv. Protein Chem., 14 (1959) 1–63.
[30] R. Wolfenden, L. Andersson, P.M. Cullis and C.C.B. Southgate, Biochemistry, 20 (1981) 849–855.
[31] J.-L. Fauchère and V. Pliska, Eur. J. Med. Chem., 18 (1983) 369–375.
[32] D.C. Rees, I. DeAntonio and D. Eisenberg, Science, 245 (1989) 510–513.
[33] B.W. Matthews, H. Nicholson and W.J. Becktel, Proc. Nat. Acad. Sci. USA, 84 (1987) 6663–6667.