J|A|C|S
ARTICLES
Published on Web 10/04/2002

Stabilizing Interactions between Aromatic and Basic Side Chains in α-Helical Peptides and Proteins. Tyrosine Effects on Helix Circular Dichroism

Charles D. Andrew,[†] Samita Bhattacharjee,[‡] Nicoleta Kokkoni,[†] Jonathan D. Hirst,[‡] Gareth R. Jones,[§] and Andrew J. Doig*,[†]

Contribution from the Department of Biomolecular Sciences, UMIST, P.O. Box 88, Manchester M60 1QD, U.K., Department of Chemistry, University of Nottingham, University Park, Nottingham NG7 2RD, U.K., and Daresbury Laboratory, Daresbury, Warrington, Cheshire WA4 4AD, U.K.

Received July 10, 2002

Abstract: Here we investigate the structures and energetics of interactions between aromatic (Phe or Tyr) and basic (Lys or Arg) amino acids in α-helices. Side chain interaction energies are measured using helical peptides, by quantifying their helicities with circular dichroism at 222 nm and interpreting the results with Lifson−Roig-based helix/coil theory. A difficulty in working with Tyr is that the aromatic ring perturbs the CD spectrum, giving an incorrect helicity. We calculated the effect of Tyr on the CD at 222 nm by deriving the intensities of the bands directly from the electronic and magnetic transition dipole moments through the rotational strengths corresponding to each excited state of the polypeptide. This gives an improved value of the helix preference of Tyr (from 0.48 to 0.35) and a correction to the helicity for the peptides containing Tyr. We find that Phe−Lys, Lys−Phe, Phe−Arg, Arg−Phe, and Tyr−Lys are all stabilizing by −0.10 to −0.18 kcal·mol^{-1} when placed *i*, *i* + 4 on the surface of a helix in aqueous solution, despite the great difference in polarity between these residues. Interactions between these side chains have previously been attributed to cation−π bonds. A survey of protein structures shows that they are in fact predominantly hydrophobic interactions between the CH_2 groups of Lys or Arg and the aromatic rings.

Introduction

α-Helices in aqueous solution adopt a large number of structures, with fully helical, fully coil, and partly helical conformations all populated. For a complete understanding of helix formation and stability, all the factors contributing to this equilibrium need to be assessed thoroughly. These include the helix-forming tendencies of constituent amino acids,[1−10] capping preferences at the carboxyl and amino termini,[11−21] and side chain interactions. Helix-stabilizing *i*, *i* + 4 side chain−side chain interactions identified previously include salt bridges,[22−30] hydrogen bonds,[29−32] and hydrophobic interactions,[33−36] but

* Corresponding author. Phone: +44-161-200 4224. Fax: +44-161-236 0409. E-mail: Andrew.Doig@umist.ac.uk.
† UMIST.
‡ University of Nottingham.
§ Daresbury Laboratory.

(1) O'Neil, K. T.; DeGrado, W. F. *Science* **1990**, *250*, 646−51.
(2) Scholtz, J. M.; Qian, H.; York, E. J.; Stewart, J. M.; Baldwin, R. L. *Biopolymers* **1991**, *31*, 1463−70.
(3) Chakrabartty, A.; Kortemme, T.; Baldwin, R. L. *Protein Sci.* **1994**, *3*, 843−52.
(4) Pace, C. N.; Scholtz, J. M. *Biophys. J.* **1998**, *75*, 422−427.
(5) Horovitz, A.; Matthews, J. M.; Fersht, A. R. *J. Mol. Biol.* **1992**, *227*, 560−568.
(6) Park, S. H.; Shalongo, W.; Stellwagen, E. *Biochemistry* **1993**, *32*, 7048−53.
(7) Blaber, M.; Zhang, X. J.; Lindstrom, J. D.; Pepiot, S. D.; Baase, W. A.; Matthews, B. W. *J. Mol. Biol.* **1994**, *235*, 600−624.
(8) Rohl, C. A.; Chakrabartty, A.; Baldwin, R. L. *Protein Sci.* **1996**, *5*, 2623−37.
(9) Myers, J. K.; Pace, C. N.; Scholtz, J. M. *Proc. Nat. Acad. Sci. U.S.A.* **1997**, *94*, 2833−2837.
(10) Yang, J.; Spek, E. J.; Gong, Y.; Zhou, H.; Kallenbach, N. R. *Protein Sci.* **1997**, *6*, 1264−72.
(11) Presta, L. G.; Rose, G. D. *Science* **1988**, *240*, 1632−41.
(12) Richardson, J. S.; Richardson, D. C. *Science* **1988**, *240*, 1648−52.
(13) Lyu, P. C.; Zhou, H. X. X.; Jelveh, N.; Wemmer, D. E.; Kallenbach, N. R. *J. Am. Chem. Soc.* **1992**, *114*, 6560−6562.
(14) Forood, B.; Feliciano, E. J.; Nambiar, K. P. *Proc. Nat. Acad. Sci. U.S.A.* **1993**, *90*, 838−42.
(15) Doig, A. J.; Chakrabartty, A.; Klingler, T. M.; Baldwin, R. L. *Biochemistry* **1994**, *33*, 3396−3403.
(16) Doig, A. J.; Baldwin, R. L. *Protein Sci.* **1995**, *4*, 1325−36.
(17) Petukhov, M.; Muñoz, V.; Yumoto, N.; Yoshikawa, S.; Serrano, L. *J. Mol. Biol.* **1998**, *278*, 279−89.
(18) Petukhov, M.; Uegaki, K.; Yumoto, N.; Yoshikawa, S.; Serrano, L. *Protein Sci.* **1999**, *8*, 2144−2150.
(19) Cochran, D. A. E.; Penel, S.; Doig, A. J. *Protein Sci.* **2001**, *10*, 463−70.
(20) Cochran, D. A. E.; Doig, A. J. *Protein Sci.* **2001**, *10*, 1305−1311.
(21) Petukhov, M.; Uegaki, K.; Yumoto, N.; Serrano, L. *Protein Sci.* **2002**, *11*, 766−777.
(22) Marqusee, S.; Baldwin, R. L. *Proc. Nat. Acad. Sci. U.S.A.* **1987**, *84*, 8898−8902.
(23) Lyu, P. C.; Marky, L. A.; Kallenbach, N. R. *J. Am. Chem. Soc.* **1989**, *111*, 2733−2734.
(24) Horovitz, A.; Serrano, L.; Avron, B.; Bycroft, M.; Fersht, A. R. *J. Mol. Biol.* **1990**, *216*, 1031−44.
(25) Gans, P. J.; Lyu, P. C.; Manning, P. C.; Woody, R. W.; Kallenbach, N. R. *Biopolymers* **1991**, *31*, 1605−1614.
(26) Merutka, G.; Stellwagen, E. *Biochemistry* **1991**, *30*, 1591−1594.
(27) Stellwagen, E.; Park, S.-H.; Shalongo, W.; Jain, A. *Biopolymers* **1992**, *32*, 1193−1200.
(28) Huyghues-Despointes, B. M.; Scholtz, J. M.; Baldwin, R. L. *Protein Sci.* **1993**, *2*, 80−5.
(29) Scholtz, J. M.; Qian, H.; Robbins, V. H.; Baldwin, R. L. *Biochemistry* **1993**, *32*, 9668−76.
(30) Huyghues-Despointes, B. M.; Baldwin, R. L. *Biochemistry* **1997**, *36*, 1965−70.
(31) Huyghues-Despointes, B. M.; Klingler, T. M.; Baldwin, R. L. *Biochemistry* **1995**, *34*, 13267−71.

10.1021/ja027629h CCC: $22.00

many have been analyzed only qualitatively. Interaction energies have been measured for Asp−Lys, Lys−Asp, Glu−His, His−Glu,[37] Lys−Lys,[27] Lys−Tyr, Tyr−Val, Tyr−Lys,[38] Gln−Asp,[31] Trp−His,[39] Glu−Lys,[27,29,37,40] Phe−Met, Met−Phe,[35,36] Lys−Glu,[29,37] Gln−Glu[29], Ile−Lys, Val−Lys, Ile−Arg,[41] and Gln−Asn.[32] Here we analyze interactions between aromatic and basic side chains.

Interactions between aromatic and basic groups have been increasingly recognized as important in proteins and generally attributed to cation−π bonds.[39,42−60] In this interaction the cationic group is positioned above the center of the aromatic ring, interacting with the delocalized electron cloud and quadrupole of the ring. The interactions we study here have previously been attributed to cation−π bonds.

Although previous investigations of aromatic−basic interactions have been carried out in helical peptides,[39,48,58,59] not all have determined free energies for the interactions. Here, we measure the free energies of Phe−Lys, Lys−Phe, Phe−Arg, Arg−Phe, and Tyr−Lys *i*, *i* + 4 pairs by placing them on the surface of alanine-based helical peptides and measuring their effect on helix content using CD. We find these interactions are stabilizing and discover a preference for certain orientations and amino acid pairs. Our crystal structure analysis indicates that the dominant effect is hydrophobic, rather than cation−π.

The characteristic CD spectrum of a protein depends on its secondary structure content;[61,62] for example, an α-helix consists of a positive band at 190 nm and two negative bands at 208 and 222 nm, whereas the β-sheet proteins have a maximum at 195 nm and a minimum in the region 210−220 nm. The helix content in peptides is thus often determined by the intensity of the 222 nm peak. However, it is widely recognized that the electronic transitions in the side chain chromophores can contribute to the spectrum in the far-UV region.[63] This can complicate the interpretation and analysis of secondary structure. Of particular importance are the residues with aromatic side chains: Tyr, His, Trp, and Phe. In peptides designed to investigate aromatic interactions, potential artifacts resulting from distortions of the CD spectrum in addition to secondary structure are unavoidable. We have therefore calculated from first principles the influence of Tyr on the circular dichroism of two alanine-based model peptides. This allows a more accurate assessment of the helix content of peptides with Tyr at interior positions. Our results suggest that the contribution of Tyr to $[\theta]_{222}$ is dependent on the precise orientation of the Tyr residue and can vary from around −4000 to +4000 deg·cm^2·dmol^{-1}. This allows more accurate measurements of the helix preference of Tyr and the strength of the Tyr−Lys interaction.

Peptide Design. Peptides are based on the control sequence Ac-AAAK*F*AAAAK*F*AAAAKAKAGY-NH_2, utilizing the high helix propensity of Ala. Ala/Lys-based sequences are monomeric in aqueous solution,[64] especially given the high frequency of positively charged groups in all our peptides. In this sequence two Phe−Lys *i*, *i* + 5 pairs are present between residues 5 and 10 and residues 11 and 16. These control peptides have the basic−aromatic residues in their pairs on opposite faces of the helix where they cannot interact. This control sequence is altered to Ac-AAAKA*F*AAAK*A*F*A*AAA*K*AKAGY-NH_2, creating two *i*, *i* + 4 pairs between the same Phe and Lys residues by interchanging the Phe at 5 with the alanine at 6 and likewise the Phe at 11 with the alanine at 12. These two nonpolar/polar *i*, *i* + 4 pairs are now on the same face of the helix in a position to interact. Sequences are designed to avoid all *i*, *i* + 3 and *i*, *i* + 4 interactions between other side chains. Gly is a helix breaker and is placed between Tyr and the rest of the sequence to prevent unwanted interaction between the aromatic chromophore and the rest of the helix, which may contribute to the circular dichroism signal.[65] Tyr is positioned at the C-terminus to allow concentration determination. The N- and C-termini are blocked with an acetyl group and an amide group, respectively, preventing destabilizing interactions with these charged termini and the helix dipole. This also decreases the number of free main chain amine and carboxyl groups, by creating an extra hydrogen bond at each terminus. Acetyl groups also substantially stabilize the helix.[15] Peptides are designed to be approximately 50% helical, where ellipticity is most sensitive to small changes in side chain interaction energy. We have investigated four possible basic interactions with Phe, namely, Phe−Lys, Lys−Phe, Phe−Arg, and Arg−Phe, to probe the effects of swapping the side chain order and substituting Lys for Arg. One Tyr interaction was studied, Tyr−Lys, in view of the difficulties in dealing with the Tyr CD effects. The peptide sequences are given in Table 1.

(32) Stapley, B. J.; Doig, A. J. *J. Mol. Biol.* **1997**, *272*, 465−473.
(33) Padmanabhan, S.; Baldwin, R. L. *Protein Sci.* **1994**, *3*, 1992−7.
(34) Padmanabhan, S.; Baldwin, R. L. *J. Mol. Biol.* **1994**, *241*, 706−13.
(35) Stapley, B. J.; Rohl, C. A.; Doig, A. J. *Protein Sci.* **1995**, *4*, 2383−91.
(36) Viguera, A. R.; Serrano, L. *Biochemistry* **1995**, *34*, 8771−9.
(37) Smith, J. S.; Scholtz, J. M. *Biochemistry* **1998**, *37*, 33−40.
(38) Shalongo, W.; Stellwagen, E. *Protein Sci.* **1995**, *4*, 1161−1166.
(39) Fernández-Recio, J.; Vásquez, A.; Civera, C.; Sevilla, P.; Sancho, J. *J. Mol. Biol.* **1997**, *267*, 184−197.
(40) Lyu, P. C.; Gans, P. J.; Kallenbach, N. R. *J. Mol. Biol.* **1992**, *223*, 343−350.
(41) Andrew, C. D.; Penel, S.; Jones, G. R.; Doig, A. J. *Proteins: Struct., Funct., Genet.* **2001**, *45*, 449−455.
(42) Wlodawer, A.; Walter, J.; Huber, R.; Sjolin, L. *J. Mol. Biol.* **1984**, *180*, 301−329.
(43) Perutz, M. F.; Fermi, G.; Abraham, D. J.; Poyart, C.; Bursaux, E. *J. Am. Chem. Soc.* **1986**, *108*, 1064−1078.
(44) Burley, S. K.; Petsko, G. A. *FEBS Lett.* **1986**, *203*, 139−143.
(45) Levitt, M.; Perutz, M. F. *J. Mol. Biol.* **1988**, *201*, 751−754.
(46) Singh, J.; Thornton, J. M. *J. Mol. Biol.* **1990**, *211*, 595−615.
(47) Loewenthal, R.; Sancho, J.; Fersht, A. R. *J. Mol. Biol.* **1992**, *224*, 759−770.
(48) Armstrong, K. M.; Fairman, R.; Baldwin, R. L. *J. Mol. Biol.* **1993**, *230*, 284−91.
(49) Mitchell, J. B.; Nandi, C. L.; MacDonald, I. K.; Thornton, J. M.; Price, S. L. *J. Mol. Biol.* **1994**, *239*, 315−331.
(50) Flocco, M. M.; Mowbray, S. L. *J. Mol. Biol.* **1994**, *235*, 709−717.
(51) Dougherty, D. A. *Science* **1996**, 163−168.
(52) Scrutton, N. S.; Raine, A. R. *Biochem. J.* **1996**, *319*, 1−8.
(53) Mitchell, J. B.; Laskowski, R. A.; Thornton, J. M. *Proteins* **1997**, *29*, 270−280.
(54) Hendlich, M. *Acta Crytallogr., Sect. D* **1998**, *54*, 1178−1182.
(55) Wouters, J. *Protein Sci.* **1998**, *7*, 2472−2475.
(56) Minoux, H.; Chipot, C. *J. Am. Chem. Soc.* **1999**, *121*, 10366−10372.
(57) Gallivan, J. P.; Dougherty, D. A. *Proc. Nat. Acad. Sci. U.S.A.* **1999**, *96*, 9459−9464.
(58) Olson, C. A.; Shi, Z.; Kallenbach, N. R. *J. Am. Chem. Soc.* **2001**, *123*, 6451−6452.
(59) Shi, Z.; Olson, C. A.; Kallenbach, N. R. *J. Am. Chem. Soc.* **2002**, *124*, 3284−3291.
(60) Burghart, T. P.; Juranic, N.; Macura, S.; Ajtai, K. *Biopolymers* **2002**, *63*, 261−272.
(61) Nakanishi, K.; Berova, N.; Woody, R. W. *Circular Dichroism Principles and Applications*; VCH: New York, 1994.
(62) Fasman, G. D. *Circular Dichroism and the Conformational Analysis of Biomolecules*; Plenum Press: New York, 1996.
(63) Woody, R. W., Dunker, A. K., Fasman, G. D., Eds. *Circular Dichroism and the Conformational Analysis of Biomacromolecules*; Plenum Press: New York, 1996; pp 109−157.
(64) Padmanabhan, S.; Marqusee, S.; Ridgeway, T.; Laue, T. M.; Baldwin, R. L. *Nature* **1990**, *344*, 268−70.
(65) Chakrabartty, A.; Kortemme, T.; Padmanabhan, S.; Baldwin, R. L. *Biochemistry* **1993**, *32*, 5560−5.

Table 1. Sequences and Helix Contents of Peptides

name	sequence	$[\theta]_{222}$[a]	% helicity[b]
FKc	Ac-AAAK**F**AAAA**KF**AAAA**K**AKAGY-NH_2	−12500	35.4
FKi	Ac-AAAKA**F**AAA**K**A**F**AAA**K**AKAGY-NH_2	−14600	41.1
KFc	Ac-AAAK**K**AAAA**FK**AAAA**F**AKAGY-NH_2	−13000	36.8
KFi	Ac-AAAKA**K**AAA**F**A**K**AAA**F**AKAGY-NH_2	−15600	43.8
FRc	Ac-YGAAK**F**AAAA**RF**AAAA**R**AKA-NH_2	−13000	37.1
FRi	Ac-YGAAKA**F**AAA**R**A**F**AAA**R**AKA-NH_2	−15700	44.4
RFc	Ac-AAAK**R**AAAA**FR**AAAA**F**AKAGY-NH_2	−13700	38.7
RFi	Ac-AAAKA**R**AAA**F**A**R**AAA**F**AKAGY-NH_2	−17000	47.6
YKc	Ac-AAAK**Y**AAAA**KY**AAAA**K**AKAGY-NH_2	−14900	41.9/43.5[c]
YKi	Ac-AAAKA**Y**AAA**K**A**Y**AAA**K**AKAGY-NH_2	−16400	46.0/48.0[c]

[a] Mean residual ellipticities measured in deg·cm^2·dmol^{-1} at 222 nm. [b] Calculated as $[\theta]_{222}$(obsd) − $[\theta]_{222}$(coil)/[−42500(1 − 3/n) − $[\theta]_{222}$(coil)], where $[\theta]_{222}$(coil) is 640 deg·cm^2·dmol^{-1} and n is the number of amino acids in the peptide.[8] [c] Correction for each Tyr residue in the helix that affects the CD signal. The helix content is thus $[\theta]_{222}$(obsd) − $[\theta]_{222}$(coil)/[−42500(1 − 3/n) − $[\theta]_{222}$(coil) + $[\theta]_{222}$(aromatic)], where $[\theta]_{222}$(coil) is 640 deg·cm^2·dmol^{-1}, $[\theta]_{222}$(aromatic) is determined from Tables 2 and 3, and n is the number of amino acids in the peptide.[8]

Methods

Application of Lifson−Roig-Based Theory. Alanine-based peptides in aqueous solution form a complex population of fully helical, fully coil, and partly helical structures. To interpret this equilibrium quantitatively, it is essential to include the structure and stability of every conformation. Lifson−Roig-based helix/coil models do this by assigning each residue to either a helix (h) or a coil (c) state. Conformations are then defined as strings of h's and c's. The contribution a residue makes to the stability of a conformation depends on its position in the helix. These include parameters for helix nucleation (v), helix interiors (w), N- and C-capping (n and c), the N-terminal helix positions N1, N2, and N3 ($n1$, $n2$, and $n3$), and side chain interactions (p and q).[8,15,35,66,67] In particular, the parameter p is the statistical weight for formation of an i, $i + 4$ side chain interaction, and the free energy of the interaction is given by $\Delta G_{i,i+4} = -RT \ln p$. For the interaction to occur, the five consecutive residues from i to $i + 4$ must be in a helical conformation. Hence, all peptide conformations that fulfill this criterion are changed in stability by $\Delta G_{i,i+4}$, thus rigorously quantifying the effect of a side chain interaction on the helix/coil equilibrium and hence on the experimentally measured mean helix content. A computer program implementing the helix/coil model, SCINT2,[8] was used to analyze the peptides. The program evaluates the partition function of any sequence given w, v, n, c, and p values for the constituent amino acids and predicts all properties of the helix/coil equilibrium, including the fractional helicity.

In predicting the helix contents of FKi, KFi, FRi, Rfi, and YKi, the only unknowns are p_{FK}, p_{KF}, p_{FR}, p_{RF}, and p_{YK}, respectively. To determine the free energy of these i, $i + 4$ interactions in an α-helix, the p value is varied by intervals of 0.1 until the calculated helix content agrees with experiment. We estimate the greatest experimental error to be in the measurement of helicity and to be about +3%, arising from pipetting errors. An estimate of the error in p was evaluated by repeating the fitting procedure using experimental helicities increased or decreased by 3%.

SCINT2 is available from the web page http://www.bi.umist.ac.uk/users/mjfajdg/HC.htm. Results using the AGADIR program,[68] which predicts the helix content of any peptide, were obtained from the web site http://www.embl-heidelberg.de/cgi/agadir-wrapper.pl.

Peptide Synthesis. Peptides were synthesized on an Applied Biosystems 431A peptide synthesizer using Fmoc solid-phase chemistry. 9-Fluoroenylmethoxycarbonyl amino acids (CN Biosciences) were coupled to rink amide resin (CN Biosciences) using 2-(1*H*-Benzotriazol-1-yl)-1,1,3,3-tetramethyluronium tetrafluoroborate (TBTU) and *N*-hydroxybenzotriazole·H_2O (HOBt) with an *N*,*N*-methylpyrrolidinone (NMP) solvent and diisopropylethylamine base. Acetylation of N-termini was carried out with pyridine and acetic anhydride. Cleavage from the resin and removal of Lys, Arg, and Tyr side chain protecting groups was accomplished with 95% trifluoroacetic acid, 2.5% triisopropylsilane, and 2.5% H_2O. Peptides were purified using C_{18} reversed-phase HPLC (Hewlett-Packard series 1100) using 5−40% H_2O/acetonitrile gradients in the presence of 0.1% trifluoroacetic acid. Peptide verification was achieved by electrospray mass spectrometry at the Michael Barber Centre for Mass Spectrometry, UMIST, Manchester, U.K.

Circular Dichroism Measurements. CD measurements were made using a Jasco J810 spectropolarimeter. Peptides were measured in 10 mM NaCl, 5 mM sodium phosphate buffer at 273 K, pH 7.0, in a 0.1 cm quartz cell. Peptide concentrations were determined by measuring the Tyr UV absorbance at 275 nm of diluted aliquots of stock solution in water or 6.9 M guanidine hydrochloride, using $\epsilon_{275} = 1390$ M^{-1} cm^{-1} and $\epsilon_{275} = 1450$ M^{-1} cm^{-1}, respectively. CD measurements are given as mean residual ellipticity at 222 nm ($[\theta]_{222}$) in units of deg·cm^2·dmol^{-1}. Helix content was calculated as $[\theta]_{222}$(obsd) − $[\theta]_{222}$(coil)/[−42500(1 − 3/n) − $[\theta]_{222}$(coil)], where $[\theta]_{222}$(coil) is 640 deg·cm^2·dmol^{-1} and n is the number of amino acids in the peptide with blocked termini.[8]

Calculations of the Effect of Tyrosine on the Circular Dichroism of Peptides. The YKc and YKi peptides were modeled as the sequences Ac-AAAAYAAAAAYAAAAAAAAAY-NH_2 and Ac-AAAAAYAA-AAAYAAAAAAAAY-NH_2. Calculations of the electronic structure of polypeptides are too demanding for fully ab initio methods. A popular technique for computing the CD spectra of polypeptides is the matrix method.[69] The intensities of the bands are derived directly from the electronic and magnetic transition dipole moments through the rotational strengths corresponding to each excited state of the polypeptide, given by the imaginary part of the product of the electronic and magnetic transition dipole moments.[70] The rotational strength $\mathbf{R}_{0i}$ of an electronic transition $i \leftarrow 0$ is given by

$$\mathbf{R}_{0i} = \mathrm{Im}(\langle\varphi_0|\mu_e|\varphi_i\rangle)\cdot(\langle\varphi_i|\mu_m|\varphi_0\rangle) \quad (1)$$

where φ_0 is the ground-state wave function, φ_i is the excited-state wave function, and μ_e and μ_m are the electronic and magnetic transition dipole moments, respectively.

The matrix method begins by considering a polypeptide as a collection of M noninteracting chromophoric groups. The excited-state wave function of the whole molecule is expressed as a linear superposition of basis functions Φ_{ia}, involving the n_i excitations within each chromophoric group:

$$\Psi_T = \sum_i^M \sum_a^{n_i} C_{ia}\Phi_{ia} \quad (2)$$

(66) Sun, J. K.; Penel, S.; Doig, A. J. *Protein Sci.* **2000**, *9*, 750−754.
(67) Doig, A. J. *Biophys. Chem.*, in press.
(68) Muñoz, V.; Serrano, L. *Nat. Struct. Biol.* **1994**, *1*, 399−409.
(69) Bayley, P. M.; Nielsen, E. B.; Schellman, J. A. *J. Phys. Chem.* **1969**, *73*, 228−243.
(70) Rosenfeld, L. *Z. Phys.* **1928**, *52*, 161−174.

Each basis function is a product of *M* monomer wave functions. The basis set is further restricted to allow only one group to be excited; electronic excitations may occur only within a group but not between the groups. Thus

$$\Phi_{ia} = \varphi_{10}\cdots\varphi_{ia}\cdots\varphi_{j0}\cdots\varphi_{M0} \tag{3}$$

where φ_{ia} represents the wave function of chromophore *i*, which has undergone an electronic excitation $a \leftarrow 0$. In general, each transition from the ground state to one of the excited states may have a nonzero rotational strength at its particular transition energy, and the CD spectrum is the sum of all these rotational strengths.

A Hamiltonian matrix is constructed: the diagonal elements are the excitation energies of the single chromophores, and the off-diagonal elements describe the interactions between different chromophoric groups. If the interactions between individual chromophoric groups are assumed to be purely electrostatic in nature, then the off-diagonal elements are computed from the electrostatic interaction between charge densities and have the form

$$V_{i0a;j0b} = \int_{\mathbf{r}_{i1}}\int_{\mathbf{r}_{j1}} \frac{\rho_{i0a}(\mathbf{r}_{i1})\,\rho_{j0b}(\mathbf{r}_{j1})}{4\pi\epsilon_0 r_{i1,j1}}\,\mathrm{d}\tau_{i1}\,\mathrm{d}\tau_{j1} \tag{4}$$

where ρ_{i0a} and ρ_{j0b} represent the permanent and transition electron densities on chromophores *i* and *j*, respectively. The matrix method employs parameters to describe the above charge distributions associated with the different electronic states of the chromophoric groups of the protein. For this study these parameters are taken from previous calculations on *N*-methylacetamide (NMA) in solution using the complete-active-space self-consistent field method implemented within a self-consistent reaction field (CASSCF/SCRF).[71−74] The interaction potentials were evaluated by representing the charge densities with a set of point charges (or monopoles) and the point charges fitted to reproduce the ab initio electrostatic potential of the various states.[75]

If only $n\pi^*$ and $\pi_{nb}\pi^*$ transitions are considered, then in the simplest case of a diamide the Hamiltonian matrix takes the form

$$H = \begin{pmatrix} E^1_{n\pi^*} & V^{11}_{n\pi^*\pi\pi^*} & V^{12}_{n\pi^*n\pi^*} & V^{12}_{n\pi^*\pi\pi^*} \\ V^{11}_{n\pi^*\pi\pi^*} & E^1_{\pi\pi^*} & V^{21}_{n\pi^*\pi\pi^*} & V^{12}_{\pi\pi^*\pi\pi^*} \\ V^{12}_{n\pi^*n\pi^*} & V^{21}_{n\pi^*\pi\pi^*} & E^2_{n\pi^*} & V^{22}_{n\pi^*\pi\pi^*} \\ V^{12}_{n\pi^*\pi\pi^*} & V^{12}_{\pi\pi^*\pi\pi^*} & V^{22}_{n\pi^*\pi\pi^*} & E^2_{\pi\pi^*} \end{pmatrix} \tag{5}$$

Diagonalization of the matrix gives the eigenvalues and eigenvectors of the transitions of the protein. The eigenvalues are the energies of the transitions of the polypeptides, and the eigenvectors are the mixing coefficients giving the contributions of the excited states of the individual groups to the delocalized excited states of the polypeptides. The eigenvectors are used to calculate the rotational strengths corresponding to each excited state of the peptide, as described in eq 1, and then the CD is calculated.

Protein Structure Generation. The macromolecular modeling software CHARMM was used.[76] The peptides were modeled with the dihedral angles $\phi = -63°$ and $\psi = -41°$ for the backbone and the dihedral angle χ_1 for the side chain Tyr residues set at one of the *trans* (*t*), *gauche*$^+$ (g^+), or *gauche*$^-$ (g^-) orientations. Thus, for the above two peptides with three side chain tyrosines each in three orientations, 27 different conformations are given for each peptide. Lowest energy conformations were obtained through the steepest descents energy minimization for these static peptide structures within the generalized Born implicit continuum solvent model[77] with dielectric constant $\epsilon = 80$ for water.

In the continuum approach, the solvent is modeled by a continuum characterized by a dielectric constant in which there is a cavity that contains the solute. This provides a macroscopic description of the solvent in which short-range solute−solvent interactions, such as hydrogen bonding, are neglected. To generate the desired geometry, a restraining potential on the dihedral angles was used with the force constant 1000 kcal·mol^{-1}·rad^{-2}.

Circular Dichroism Calculations. As described above, the calculations of CD are based on the matrix method. To describe the electronic properties of the peptide chromophores, we have employed a recent parameter set, which has considerably improved the accuracy of protein circular dichroism calculations.[78] Two peptide transitions are considered, the $n\pi^*$ transition at 222 nm and the $\pi_{nb}\pi^*$ transition at 193 nm. Higher energy transitions are not considered, as they lie outside the region of interest and the reliability of the parameter sets describing them has yet to be fully established. In addition, four transitions due to the Tyr chromophores are included, L_a at 227 nm, L_b at 278 nm, B_b at 193 nm, and B_a at 192 nm. We use the recent parameter set for Tyr of Woody and co-workers.[79,80] The other methodological details are the same as those in the earlier theoretical study on proteins.[78] Transitions in the peptide circular dichroism spectrum are assumed to have a bandwidth of 15.5 nm.

In this study, we are concerned with the helicity of the peptides. We focus, therefore, on the calculated intensity at one wavelength of our computed spectra, 222 nm. The intensity at 222 nm, $[\theta]_{222}$, is widely used to estimate the helicity of proteins and peptides. To investigate the influence of the Tyr residue on the circular dichroism, we compare calculations of the circular dichroism spectra with and without parameters for the transitions associated with the Tyr.

Crystal Structure Survey. We surveyed 299 protein structures[81] in the PDB, with a resolution of 2.0 Å or better, with an *R* factor lower than 20%, and with less than 25% identity for the occurrence of certain *i*, $i + 4$ pairs. We used the program SSTRUC[82] to find the secondary structure of each protein sequence. From the 1531 helices isolated, we obtained the amino acid frequencies and their χ_1, χ_2, χ_3, and χ_4 angles. In the following helical sequence, terminal residues N1, N2, and N3 and C1, C2, and C3 have unsatisfied main chain amine and carboxyl groups, respectively:

Ncap-N1-N2-N3-XXXXX-C3-C2-C1-Ccap

These residues have unique structural properties, so *i*, $i + 4$ pairs present within these regions were not considered. Only *i*, $i + 4$ pairs containing both residues in a helical region (X) of five residues or greater were considered as all residues in this stretch have both amine and carboxyl groups satisfied by backbone helical hydrogen bonds. With 6 terminal residues and a minimum of 5 central helical residues required for an *i*, $i + 4$ pair to be considered, only helices of 11 residues or greater were examined.

A window of five residues was moved along the helices, and any pair of residues of interest spaced *i*, $i + 4$, with themselves and all intervening residues falling within the helical X region, was considered an occurrence. A total of 5546 *i*, $i + 4$ motifs were analyzed for specific pairs. The expected number of Y−Z pairs is calculated as the product of the probability of a Y residue in the X helical region and the probability of a Z residue in the X helical region, divided by the number of *i*, $i + 4$ motifs. Side chain χ_1 angles are classified into three

(71) Karlström, G. *J. Phys. Chem.* **1988**, *92*, 1315−1318.
(72) Karlström, G. *J. Phys. Chem.* **1989**, *93*, 4952−4955.
(73) Bernhardsson, A.; Lindh, R.; Karlstrom, G.; Roos, B. O. *Chem. Phys. Lett.* **1996**, *151*, 141−149.
(74) Serrano-Andrés, L.; Fülscher, M. P.; Karlstrom, G. *Int. J. Quantum Chem.* **1997**, *65*, 167−181.
(75) Besley, N. A.; Hirst, J. A. *J. Phys. Chem. A* **1998**, *102*, 10791−10797.
(76) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.
(77) Dominy, B. N.; Brooks, C. L. *J. Phys. Chem. B* **1999**, *103*, 3765−3773.
(78) Besley, N. A.; Hirst, J. A. *J. Am. Chem. Soc.* **1999**, *121*, 9636−9644.
(79) Woody, R. W.; Sreerama, N. *J. Chem. Phys.* **1999**, *111*, 2844−2845.
(80) Sreerama, N.; Manning, M. C.; Powers, M. E.; Zhang, J.-X.; Goldenberg, D. P.; Woody, R. W. *Biochemistry* **1999**, *38*, 10814−10822.
(81) Penel, S.; Hughes, E.; Doig, A. J. *J. Mol. Biol.* **1999**, *287*, 127−143.
(82) Smith, D. K.; Thornton, J. M. SSTRUC program; Department of Biochemistry and Molecular Biology, University College London, 1989.

Table 2. CD Calculations in Ac-AAAAYAAAAAYAAAAAAAAAY-NH_2 (YKc Model)

side chain conformation			$[\theta]_{222}$(calcd) (deg·cm²·dmol⁻¹)			
Tyr 1	Tyr 2	Tyr 3	without Tyr	with Tyr	difference	difference × population[a]
t	t	t	−22653	−19164	3489	1003
t	t	g^+	−22755	−21967	788	113
t	t	g^-	−22714	−20139	2575	11
t	g^+	t	−22821	−19482	3339	480
t	g^-	t	−22563	−19175	3388	15
t	g^+	g^+	−22678	−22314	364	26
t	g^+	g^-	−22569	−20078	1491	3
t	g^-	g^+	−22857	−22445	412	1
t	g^-	g^-	−22473	−19674	2799	0
g^+	t	t	−23164	−24735	−1571	−226
g^-	t	t	−22699	−20665	2034	9
g^+	t	g^+	−23280	−24272	992	−71
g^+	t	g^-	−23062	−21883	1179	3
g^-	t	g^+	−22683	−24576	−1893	−4
g^-	t	g^-	−22523	−21905	618	0
g^+	g^+	t	−24037	−24304	−267	−19
g^+	g^-	t	−23370	−25250	−1880	−4
g^-	g^+	t	−22512	−25276	−2764	−6
g^-	g^-	t	−20834	−19790	1045	0
g^+	g^+	g^+	−25021	−24076	945	34
g^+	g^+	g^-	−24488	−22945	1543	2
g^+	g^-	g^+	−23566	−25263	−1697	−2
g^+	g^-	g^-	−23333	−23064	269	0
g^-	g^+	g^+	−22528	−23575	−1047	−1
g^-	g^+	g^-	−22350	−21246	1104	0
g^-	g^-	g^+	−19799	−23712	−3913	0
g^-	g^-	g^-	−19109	−21544	−2435	0
						sum = 1366

[a] Population = product of rotamer populations (*t*, 66%; g^+, 33%; g^-, 1%[81]).

Table 3. CD Calculations in Ac-AAAAAYAAAAAYAAAAAAAAY-NH_2 (YKi Model)

side chain conformation			$[\theta]_{222}$(calcd) (deg·cm²·dmol⁻¹)			
Tyr 1	Tyr 2	Tyr 3	without Tyr	with Tyr	difference	difference × population[a]
t	t	t	−22568	−19258	3310	952
t	t	g^+	−22653	−22002	650	93
t	t	g^-	−22464	−19130	3333	15
t	g^+	t	−22566	−19299	3268	470
t	g^-	t	−22618	−19344	3274	14
t	g^+	g^+	−22355	−21941	414	30
t	g^+	g^-	−22474	−20346	2128	5
t	g^-	g^+	−22124	−20249	1875	4
t	g^-	g^-	−22043	−19419	2624	0
g^+	t	t	−23210	−23911	−700	−101
g^-	t	t	−22398	−19055	3343	15
g^+	t	g^+	−23336	−23401	−65	−5
g^+	t	g^-	−23114	−21020	2093	5
g^-	t	g^+	−22596	−24293	−1697	−4
g^-	t	g^-	−22504	−21087	1416	0
g^+	g^+	t	−24131	−23748	383	28
g^+	g^-	t	−23463	−24884	−1421	−3
g^-	g^+	t	−22708	−24923	−2215	−5
g^-	g^-	t	−20218	−17948	2270	0
g^+	g^+	g^+	−25070	−23494	1576	57
g^+	g^+	g^-	−24562	−22535	2027	2
g^+	g^-	g^+	−23691	−24710	−1018	−1
g^+	g^-	g^-	−23441	−23144	297	0
g^-	g^+	g^+	−22738	−23017	−280	0
g^-	g^+	g^-	−22551	−20819	1732	0
g^-	g^-	g^+	−17026	−18484	−1458	0
g^-	g^-	g^-	−20723	−20946	−223	0
						sum = 1570

[a] Population = product of rotamer populations (*t*, 66%, g^+, 33%, g^-, 1%[81]).

rotamers: $0° < \chi_1 \leq 120°$, *gauche*$^-$; $120° < \chi_1 \leq -120°$, *trans*; $-120° < \chi_1 \leq 0°$, *gauche*$^+$.

Results

Helix Contents of Peptides. Table 1 shows the sequences, mean residual ellipticities, and helix contents of the synthesized peptides. The control peptides FKc, KFc, FRc, RFc, and YKc with the basic−aromatic pairs spaced *i*, *i* + 5 have notably lower helix contents than the interaction peptides FKi, KFi, FRi, RFi, and YKi with the aromatic−basic pairs spaced *i*, *i* + 4. The average change in ellipticity for the three pairs ($\Delta([\theta]_{222}) = -2440$ deg·cm²·dmol⁻¹) is outside the range of experimental error (±1000 deg·cm²·dmol⁻¹). These results show qualitatively that Phe−Lys, Lys−Phe, Phe−Arg, Arg−Phe, and Tyr−Lys interactions stabilize isolated α-helices in water.

CD Calculations. Tables 2 and 3 give the effects of the Tyr side chains on the CD signal at 222 nm for the YKc and YKi peptides, modeled as 100% helical, as a function of the χ_1 rotamers of the three Tyr side chains. Effects of up to 4000 deg·cm²·dmol⁻¹ are seen, supporting the view that aromatic residues can have a significant effect on the CD. The total effect can be determined by weighting each correction by their rotamer populations. In interior helix positions the χ_1 populations are (*t*) 66%, (g^+) 33%, and (g^-) 1%.[81] Multiplying these weights by the correction for each conformation and summing give the CD correction of YKc as −1366 deg·cm²·dmol⁻¹ and the CD correction of YKi as −1570 deg·cm²·dmol⁻¹. The $[\theta]_{222}$ value for 100% helix content is given by $-42500(1 - 3/n) - [\theta]_{222}$(coil).[8] Adding the calculated corrections gives $[\theta]_{222}$ values for 100% helix content of −37155 deg·cm²·dmol⁻¹ for YKc and −35499 deg·cm²·dmol⁻¹ for YKi. The measured helix contents for YKc and YKi are thus adjusted slightly from 41.9% to 43.5% and from 46.0% to 48.0%. The theoretical estimates perhaps overestimate the contribution from the terminal Tyr, which experimentally is conformationally labile as a result of the preceding Gly. In any case, the Tyr correction is small, given that the experimental error is ±3%.

Application of Lifson−Roig-Based Theory. The experimental CD results were analyzed with SCINT2,[8] allowing free energies of the interactions to be quantitatively determined. Table 4 shows the experimental helicities of the peptides and their corresponding theoretical helicities as predicted by the algorithm, assuming no side chain interactions are present. The four experimental control peptide helicities involving Phe agree with their theoretical helicities to within experimental error, confirming that the previously measured helix/coil parameters and helix/coil theory accurately reproduce the experimental data. All five interaction peptides have experimental helicities greater than those of their controls, where no *i*, *i* + 4 interaction is present. In addition, the Phe peptides are on average 7% more helical than the predictions (i.e., where $p_{FK} = p_{KF} = p_{FR} = p_{RF} = p_{YK} = 1$), indicating a stabilizing side chain interaction.

The pairs containing a basic−aromatic *i*, *i* + 4 pair in the N → C direction have larger helicities compared to their predictions than the reverse *i*, *i* + 4 pair. For example, the FRi peptide has a helicity 10% greater than its prediction compared to the RFi peptide, which has a helicity 5% greater, indicating orientation specificity for the aromatic−polar pairs.

The *p* values for the Phe−Lys, Lys−Phe, Phe−Arg, and Arg−Phe interactions were determined by adjusting the *p* values

Table 4. Comparison of Experimental and Theoretical Helicities, AGADIR Helix Content Predictions, and Resulting Free Energies

name	exptl helicity[a]	theoretical helicity ($p = 1$)	p value to fit experiment	ΔG (kcal·mol^{-1})	ΔG range (kcal·mol^{-1})	AGADIR helix content prediction (%)[b]
FKc	35.4	35.6	N/A	N/A	N/A	45.9
FKi	41.1	34.6	1.3	−0.14	−0.05 to −0.18	57.0
KFc	36.8	38.7	N/A	N/A	N/A	34.8
KFi	43.8	38.8	1.2	−0.10	−0.05 to −0.14	51.7
FRc	37.1	35.2	N/A	N/A	N/A	49.2
FRi	44.4	34.5	1.4	−0.18	−0.14 to −0.25	60.2
RFc	38.7	42.5	N/A	N/A	N/A	41.9
RFi	47.6	42.5	1.2	−0.10	−0.05 to −0.18	56.6
YKc	43.5	51.8[c]/43.5[d]	N/A	N/A	N/A	48.4
YKi	48.0	51.2[c]/45.2[d]	1.2	−0.10	−0.05 to −0.17	59.5

[a] Table 1. [b] Reference 84. [c] With original value of w(Tyr) = 0.48. [d] With new value of w(Tyr) = 0.35.

until the calculated helix content was in agreement with experiment. The Lys−Phe and Arg−Phe side chain interactions both had a p value of 1.2, the Phe−Lys interaction had a p value of 1.3, and the Phe−Arg interaction had a p value of 1.4. As p is the equilibrium constant for the formation of an i, $i + 4$ interaction in an α-helix, the free energy for this interaction can be calculated from $\Delta G_{i,i+4} = -RT \ln p$. The Lys−Phe and Arg−Phe p values correspond to free energies of −0.10 kcal·mol^{-1}, the Phe−Lys p value corresponds to −0.14 kcal·mol^{-1}, and that for Phe−Arg corresponds to −0.18 kcal·mol^{-1}. An estimation of the error in ΔG can be calculated by allowing the experimental helicities to vary by ±3% and recalculating the p value (Table 4). This shows that all the Phe interactions are stabilizing within experimental error.

The experimental helix content of the control peptide YKc (43.5%) does not agree with the theoretical helix content (51.8%). This is probably because the helix/coil parameters for Tyr were determined without a rigorous calculation of the Tyr effect on the CD spectrum. It is therefore likely that the helix interior preference for Tyr (w(Tyr)) is in error, but the data for the control peptide YKc now allow an improved determination of this value. If w(Tyr) is changed from 0.48 to 0.35, the calculated helix content of YKc is in agreement with experiment and is brought closer to the w value of Phe (0.27[8]). If this improved value of w(Tyr) is used in the calculations for YKi, we find ΔG for the Tyr−Lys interaction to be −0.1 kcal·mol^{-1}, in excellent agreement with the other side chain interaction energies, notably that of Phe−Lys. The extra oxygen atom in Tyr compared to Phe therefore makes no significant difference in the interaction energy.

The AGADIR program[83] can also be used to extract side chain interaction energies from experimental data. We have not used AGADIR here as it overestimates the helicity of some of the control peptides such as FKc and FRc by as much as 12%. Although AGADIR shows each of the Phe−Lys, Lys−Phe, Phe−Arg, Arg−Phe, and Tyr−Lys i, $i + 4$ pairs is stabilizing, the difference in helix content between the control and interaction peptides predicted by AGADIR is much greater than in our experimental results. It is not clear why the predictions for the two algorithms should vary, as both are based on the Lifson−Roig helix/coil model. The numerous differences between the two models include the partition functions, as AGADIR uses the single-sequence approximation, the minimal helix length in AGADIR being four residues, rather than three, different parameter sets, separation of the entropic cost of fixing a residue in a helical conformation from the backbone hydrogen bond energy in AGADIR, and inclusion of a helix dipole energy and C-capping motifs in AGADIR.[67,84]

Table 5. Propensities of Phe−Lys, Lys−Phe, Phe−Arg, Arg−Phe, and Tyr−Lys Pairs in Protein Crystal Structures

pair	no. obsd[a]	no. expected[b]	propensity[c]
Phe−Lys i, $i + 4$	11	15.9 (1.3)	0.7 (0)[c]
Phe−Lys i, $i + 7$	4	8.3 (0.7)	0.5 (0)
Lys−Phe i, $i + 4$	12	15.9 (1.3)	0.8 (0)
Lys−Phe i, $i + 7$	6	8.3 (0.7)	0.7 (0)
Phe−Arg i, $i + 4$	9	13.5 (1.1)	0.7 (0.1)
Phe−Arg i, $i + 7$	5	7.0 (0.6)	0.7 (0.1)
Arg−Phe i, $i + 4$	13	13.5 (1.1)	1.0 (0.1)
Arg−Phe i, $i + 7$	9	7.0 (0.6)	1.3 (0.2)
Tyr−Lys i, $i + 4$	13	14.3 (1.2)	0.9 (0.1)
Tyr−Lys i, $i + 7$	5	7.4 (0.6)	0.7 (0)
Lys−Tyr i, $i + 4$	13	14.3 (1.2)	0.9 (0.1)
Lys−Tyr i, $i + 7$	8	7.4 (0.6)	1.1 (0.1)

[a] See the Materials and Methods, Crystal Structure Survey. [b] For the pair AB of amino acids A and B, prob(A) is the probability of A, prob(B) is the probability of B, N is the number of amino acids (prob(A) is the number of A's/N), and N_{motifs} is the number of motifs (for i, $i + 4$ pairs, the number of four consecutive residues between N4 and C4). Error on A, $\sigma(A) = [(\text{prob}(A))(1 - \text{prob}(A))/N]^{1/2}$. Error on B, $\sigma(B) = [(\text{prob}(B))(1 - \text{prob}(B))/N]^{1/2}$. Error on expected AB, $\sigma_{expected}(AB) = (\sigma(A)\ \text{prob}(B) + \sigma(B)\ \text{prob}(A))N_{motifs}$. [c] Calculated as no. obsd/no. expected.

Previous estimates for the free energies of the interactions using AGADIR were 0 kcal·mol^{-1} for Phe−Lys, Phe−Arg, and Tyr−Lys and −0.2 kcal·mol^{-1} for Lys−Phe and Arg−Phe.[83] A later version of AGADIR used an energy of −0.20 kcal·mol^{-1} for Tyr−Lys.[85] These are in good agreement with our results. Fisinger et al. calculated the energies of all side chain interactions in α-helices.[86] Their energies for Phe−Lys, Arg−Phe, and Tyr−Lys are in good agreement with ours, but those for Lys−Phe (0.4 kcal·mol^{-1}) and Phe−Arg (−0.4 kcal·mol^{-1}) are significantly different.

Crystal Structure Data. Our survey of helices in protein crystal structures provides data regarding the propensities of Phe−Lys, Lys−Phe, Phe−Arg, Arg−Phe, Tyr−Lys, and Lys−Tyr basic−aromatic i, $i + 4$ pairs (Table 5). Due to the low occurrence of these pairs, definite trends cannot be determined on the basis of their propensities. On analysis of the 12 different i, $i + 4$ and i, $i + 7$ basic−aromatic pairs, the highest occurrence in any one category is 13 pairs. A propensity of 1 means that a specific pair is observed as many times as expected, on the basis of the frequencies of the amino acids in helices. In an i, $i + 4$ pair, both residues will be on the same face of the helix.

(83) Muñoz, V.; Serrano, L. *Proteins* **1994**, *20*, 301−11.
(84) Lacroix, E.; Viguera, A. R.; Serrano, L. *J. Mol. Biol.* **1998**, *284*, 173−91.
(85) Muñoz, V.; Serrano, L. *J. Mol. Biol.* **1995**, *245*, 275−96.
(86) Fisinger, S.; Serrano, L.; Lacroix, E. *Protein Sci.* **2001**, *10*, 809−818.

Table 6. Comparison of χ_1 and χ_2 Rotamer Populations of Phe, Tyr, Lys, and Arg in *i*, *i* + 4 Pairs and in All Helices

residue	no. obsd	no. of g^-,g^- rotamers	no. of g^-,t rotamers	no. of g^-,g^+ rotamers	no. of t,g^- rotamers	no. of t,t rotamers	no. of t,g^+ rotamers	no. of g^+,g^- rotamers	no. of g^+,t rotamers	no. of g^+,g^+ rotamers
Phe[a,b]	491	2 (1%)				328 (67%)				161 (32%)
Tyr[a]	441	4 (1%)				292 (66%)				145 (33%)
Lys[a]	729	6 (1%)	6 (1%)	0 (0%)	81 (11%)	239 (33%)	19 (3%)	22 (3)	297 (41%)	59 (8%)
Arg[a]	617	1 (0.2%)	10 (2%)	0 (0%)	60 (10%)	219 (36%)	13 (2%)	5 (1%)	273 (44%)	36 (6%)
Phe−Lys[c]	11	0 (0%)				8 (73%)				3 (27%)
Phe−**Lys**[c]	11	2 (18%)	0 (0%)	0 (0%)	0 (0%)	5 (46%)	0 (0%)	1 (9%)	2 (18%)	1 (9%)
Lys−Phe[b,c]	12	0 (0%)	0 (0%)	0 (0%)	0 (0%)	2 (17%)	0 (0%)	2 (17%)	6 (50%)	2 (17%)
Lys−**Phe**[c]	12	0 (0%)				11 (92%)				1 (8%)
Phe−Arg[c]	9	0 (0%)				5 (56%)				4 (44%)
Phe−**Arg**[c]	9	0 (0%)	0 (0%)	0 (0%)	1 (11%)	5 (56%)	0 (0%)	0 (0%)	2 (22%)	1 (11%)
Arg−Phe[b,c]	13	0 (0%)	2 (15%)	0 (0%)	1 (8%)	4 (31%)	0 (0%)	0 (0%)	5 (39%)	1 (8%)
Arg−**Phe**[c]	13	0 (0%)				8 (62%)				5 (38%)
Tyr−Lys[c]	13	0 (0%)				12 (92%)				1 (8%)
Tyr−**Lys**[c]	13	0 (0%)	0 (0%)	0 (0%)	1 (8%)	6 (46%)	0 (0%)	0 (0%)	2 (15%)	4 (31%)

[a] Rotamers for residue in all helices. [b] Val has no χ_2 rotamer, g^-,g^- corresponds to χ_1 *gauche*$^-$, t,t corresponds to χ_1 *trans*, and g^+,g^+ corresponds to χ_1 *gauche*$^+$. Side chain χ_1 angles for valine are classified as $0° < \chi_1 \leq 120°$, *trans*; $120° < \chi_1 \leq -120°$, *gauche*$^+$; and $-120° < \chi_1 \leq 0°$, *gauche*$^-$. [c] Results for residue in bold.

Due to the amphiphilic nature of the helix, basic−aromatic pairs with the hydrophobic aromatics and polar Lys/Arg would be expected to have low propensities. If we select amino acids as strongly polar (Lys, Arg, His, Glu, Asp, Gln, Asn, Ser, and Thr) or strongly nonpolar (Leu, Ile, Met, Phe, and Val), the mean propensities of their pairs are (polar−polar) 1.36, (polar−nonpolar) 0.92, (nonpolar−polar) 0.82, and (nonpolar−nonpolar) 1.18.[41] The Phe−Lys and Phe−Arg *i*, *i* + 4 propensities are marginally smaller than the mean nonpolar−polar propensity, giving little indication of whether the interaction is favored.

The Phe−Lys, Phe−Arg, and Tyr−Lys *i*, *i* + 4 pairs have equal or higher propensities than the same pairs with an *i*, *i* + 7 spacing (Table 5), where the residues are on the same face of the helix as an *i*, *i* + 4 pair, but one turn further apart. This lower result for the *i*, *i* + 7 pair shows the true amphiphilic nature of the helix, as the side chains do not interact when spaced across two turns of the helix. Therefore, the higher value for the *i*, *i* + 4 basic−aromatic pair and the difference between it and the *i*, *i* + 7 propensity probably reflect the stabilizing interaction of the basic−aromatic pair in the former spacing, although there are a very small number of *i*, *i* + 7 pairs observed. If the order of the *i*, *i* + 4 pairs is reversed, their propensities are either equal to or higher than those of the pairs in the original order, indicating that a preference for this basic−aromatic order is present, which disagrees with our peptide helicities. This is most clearly seen with Arg−Phe and Phe−Arg in *i*, *i* + 4 pairs, having propensities of 1.0 and 0.7, respectively (Table 5).

In two cases though, the reverse *i*, *i* + 4 pair is lower than the reverse *i*, *i* + 7 pair, suggesting that the reverse *i*, *i* + 4 pair is not preferred. While the propensity data are suggestive, they cannot substitute for an experimental measurement of an interaction energy. Our experimental studies show that the basic−aromatic orientation Phe−Lys is more stabilizing and preferred over the reverse orientation Lys−Phe.

Inspection of rotamer populations of the individual residues in all protein helices surveyed, and the same residues when in a specific *i*, *i* + 4 pair, gives some clear trends, although only a small number of pairs are available (Table 6). For the aromatic residue in Phe−Lys, Lys−Phe, Phe−Arg, Arg−Phe, and Tyr−Lys pairs, it is seen that the preferred rotamer in the pair is already the preferred rotamer for that residue in all helices in the proteins surveyed. All of the Phe and Tyr residues in the pairs and in the proteins prefer the χ_1 *trans* conformation. Phenylalanine is found more often in the *trans* conformation when in Lys−Phe and Arg−Phe pairs than when in the reverse pairs. For the polar *i* + 4 residue in Phe−Lys, Phe−Arg, and Tyr−Lys, the preferred rotamer in pairs is χ_1 *trans*,χ_2 *trans*, although the preferred rotamer in all helices is χ_1 *gauche*$^+$,χ_2 *trans*. When considering the reverse interaction, both polar residues in the Lys−Phe and Arg−Phe pairs are *gauche*$^+$,*trans*, the preferred rotamer in all helices.

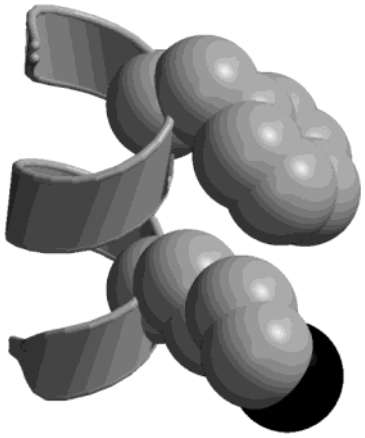

Figure 1. Phe−Lys *i*, *i* + 4 interaction between *trans*-Phe 702 and *trans*,*trans*-Lys 706, in xylanase, from the file 1xyx.pdb.

Inspection of all the crystal structures shows that Phe−Lys, Phe−Arg, and Tyr−Lys have the closest contacts, significantly better than those of the reverse Lys−Phe and Arg−Phe pairs. Within the 13 *i*, *i* + 4 Phe−Lys pairs the best contact occurs with Phe in the *trans* conformation and Lys in the *trans*,*trans* conformation. These interactions are mostly hydrophobic between the lysine carbon chain and phenyalanine aromatic ring. Figure 1 shows the hydrophobic interaction between *trans*-Phe 702 and *trans*,*trans*-Lys 706 in xylanase from the file 1xyz.pdb, which make contact via the Cϵ, CZ and Cϵ, CZ, Cδ atoms, respectively. We do not see any evidence of cation−π interactions in these structures. The charged amine group of the Lys extends beyond the Phe into the solvent. In contrast there are no close contacts between the 12 *i*, *i* + 4 Lys−Phe pairs. Even though both residues in the pair are in the most populated rotamers found for Lys (*gauche*$^+$,*trans*) and Phe (*trans*) in protein helices, an interaction does not form. The closest

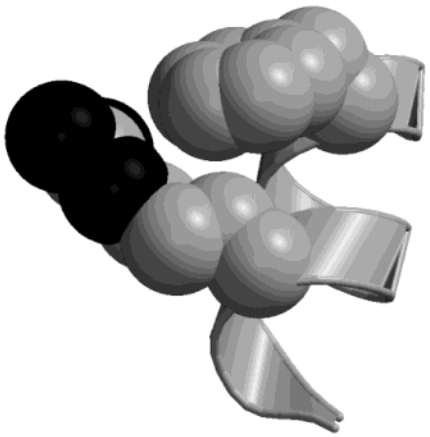

Figure 2. Phe−Arg *i*, *i* + 4 interaction between *gauche*$^+$-Phe 52 and *trans*,*trans*-Lys 56, in lysin, from the file 1lis.pdb.

distances between the residues are with a Lys residue in the *trans*,*trans* conformation. Therefore, both our peptide and crystal structure data point to the Phe−Lys orientation being preferred over Lys−Phe.

A similar but less pronounced effect is seen in the 9 Phe−Arg and 13 Arg−Phe pairs. The best contacts in the Phe−Arg pair are again seen with Phe in the *trans* conformation and Lys in the *trans*,*trans* and *trans*,*gauche*$^-$ conformations. Other close contacts occur with Lys *gauche*$^+$,*trans* and *gauche*$^+$,*gauche*$^+$. Like Phe−Lys the interactions appear to be hydrophobic between Phe and the Arg carbon chain, although the Arg nitrogens are in closer proximity to the ring than those of lysine. This latter observation raises the possibility of cation−π interactions, although hydrophobic interactions predominate in our study. Figure 2 shows the hydrophobic interaction between *gauche*$^+$-Phe 52 and *trans*,*trans*-Arg 56 in lysin from the file 1lis.pdb, which make contact predominantly via the Cϵ, CZ and Cβ, Cγ, CZ atoms, respectively.

The Arg−Phe pairs do not show as many close contacts as the Phe−Arg pairs. Although the majority of the pairs have Phe *trans* and Arg *gauche*$^+$,*trans*, the preferred rotamers in all helices, they do not form an interaction. The reasonably close contacts are seen with Phe *gauche*$^+$ and Arg *trans*,*trans*. These observations support our larger energy for the Phe−Arg over the Arg−Phe *i*, *i* + 4 pair.

On inspection of the 13 Tyr−Lys pairs, we find almost all the pairs have a reasonable amount of contact and therefore a substantially shorter average distance between the Tyr and Lys, compared with the other 4 pairs investigated. The majority of the best contacts are seen with the Tyr *trans* and the Lys *gauche*$^+$,*gauche*$^+$, although a smaller amount of good contacts are seen with Lys *gauche*$^+$,*trans*. The preference for Lys *gauche*$^+$,*gauche*$^+$ is in contrast to Lys in all helices, where only 8% of the residues adopt this conformation. Unlike the Phe−Lys and Phe−Arg pairs, when Lys is *trans*,*trans*, not as many good contacts occur. Like the other two pairs though, the interactions are again hydrophobic and seem to show a greater amount of contact. Figure 3 shows the hydrophobic interaction between *trans*-Tyr 175 and *gauche*$^+$,*trans*-Lys 179 in guanylate kinase from the file 1gky.pdb, which makes hydrophobic contact via the Cδ, Cϵ, CZ and Cδ, Cϵ atoms, respectively.

There is also a tendency for the charged Lys NH group to position itself close to the Tyr OH group with the possibility of

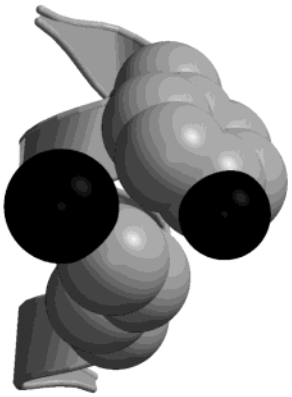

Figure 3. Tyr−Lys *i*, *i* + 4 interaction between *trans*-Phe 175 and *gauche*$^+$,*trans*-Lys 179, in guanylate kinase, from the file 1gky.pdb.

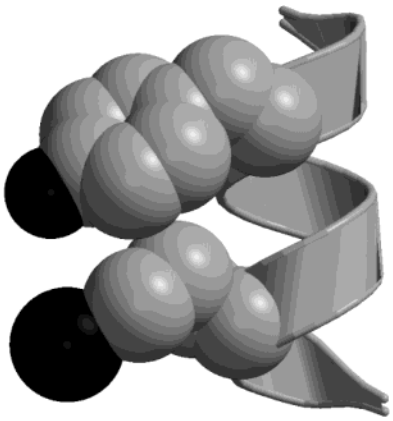

Figure 4. Tyr−Lys *i*, *i* + 4 interaction between *trans*-Phe 473 and *gauche*$^+$,*gauche*$^+$-Lys 477, in myosin, from the file 1vom.pdb.

hydrogen bond formation, although both groups could be in contact with the solvent. Figure 4 shows the hydrophobic interaction and possible hydrogen bond between *trans*-Tyr 473 and *gauche*$^+$,*trans*-Lys 477 in guanylate kinase from the file 1vom.pdb, which make contact via the Cδ, Cϵ and Cγ, Cδ atoms, respectively.

Discussion

We have quantitatively measured the energetics of the Phe−Lys, Lys−Phe, Phe−Arg, Arg−Phe, and Tyr−Lys basic−aromatic *i*, *i* + 4 interactions in an isolated peptide helix, with the interaction separated from all other stabilizing factors using helix/coil theory and isomeric control peptides. The results (Tables 1 and 4) clearly show that each of the five pairs interact favorably when placed *i*, *i* + 4 in a helix, with a preference for the aromatic−polar orientation. Free energies calculated from modified Lifson−Roig theory (Table 4) produce values between −0.10 and −0.18 kcal·mol^{-1}. Previous studies by Olson, Shi, and Kallenbach of basic−aromatic interactions involving Phe and Arg spaced *i*, *i* + 4 in helical peptides conclude that the interaction is not stabilizing, although NMR data show the residues are close enough to make contact.[58,59] A possible explanation for the same helicity in their control and test peptides could be linked to peptide design. In their study the C-terminal-positioned Arg is moved one place closer to the N-terminus, which may lower the helix content as favorable positive Arg with negative C-terminus interactions are affected. This destabilizing effect is counterbalanced by the introduction of a stabilizing Phe−Arg *i*, *i* + 4 interaction. Our studies avoid this

issue by moving the aromatic groups only, revealing the effect of the stabilizing interaction.

In agreement with our previous work on stabilizing interactions between nonpolar and polar side chains,[41] the Phe−Lys, Lys−Phe, Phe−Arg, Arg−Phe, and Tyr−Lys energies do not agree with the simple view of charged and hydrophobic amino acids being unable to interact favorably with one another. Our crystal structure examinations also add more weight to the dual polar and hydrophobic character of Lys and Arg. In both these interactions the polar groups are not involved and are free to hydrogen bond with other groups or water. We find little evidence of stacked or parallel cation−π interactions in our study (Figures 1−4). The stabilizing effects result from hydrophobic contacts between the aromatic rings and the CH_2 groups in the Lys and Arg side chains. Stabilizing interactions between nonpolar and polar side chains have been proposed in the context of binding between sense and antisense peptides, including Phe−Lys bonds in particular.[87−89] The crystal structure analysis must be treated with some caution, however. The number of cases present is not high, and additional contacts present in the protein may affect the interactions.

(87) Brentani, R. R. *J. Theor. Biol.* **1988**, *135*, 495−499.
(88) Chaiken, I. *J. Chromatogr.* **1992**, *597*, 29−36.
(89) Heal, J. R.; Roberts, G. W.; Raynes, J. G.; Bhakoo, A.; Miller, A. D. *ChemBioChem* **2002**, *3*, 136−151.

The CD calculations show that the proximity of the aromatic side chain ring to the backbone carbonyl groups can cause the coupling of the transitions of the carbonyl groups with the $\pi_{nb}\pi^*$ transition of the aromatic group. The estimates of helicity derived from the intensity at 222 nm for Tyr-containing peptides will tend to underestimate the helical content of the peptides, though by less than previously thought.[65] Consideration of the Tyr effect allows an improved estimate of the helix preference of Tyr and shows that the Tyr−Lys interaction is not significantly different from the Phe−Lys interaction. The Tyr effect on the CD is strongly dependent on the side chain conformation.

Acknowledgment. We thank Ian Fleet and Simon Gaskell at the Simon Barber Centre for Mass Spectrometry, UMIST, for peptide verification. C.D.A. is the grateful recipient of a BBSRC (U.K.) studentship with a CASE award (Daresbury). Financial support from the BBSRC (Grant No. 42/B15240) is acknowledged. We thank the Wellcome trust for an equipment grant for the CD spectrometer (reference 057318) and Neville Kallenbach for useful discussions.

JA027629H