7 Concluding Remarks

The aim of the presented thesis was to investigate the interactions of aromatic groups and their role in the processes of chemical and biological recognition. In particular, we investigated the role of aromatic residues in the protein hydrophobic core for the stability of the protein native fold. We showed that the interactions of aromatic residues not only with other aromatic and aliphatic sidechains but also with the peptide bond portion of the protein backbone can provide a substantial contribution to the protein stabilisation. Two more specific examples of the aromatic interactions, the interaction of aromatic residue with proline and the aromatic–peptide bond interaction were both examined in more detail.

The chapter on the IEM focused on aromatic interactions from the perspective of all the pairwise residue interactions in protein. The aromatic residues were shown to play a key role in the stabilisation of the Trp-cage model protein. It was also shown that the IEM can be used as a reliable method for the identification of the energetically most stabilising residues. Several computational methods were tested and it was shown that when used with the Amber parm94 force field and GB solvent model, the IEM gives satisfactory results while being computationally feasible for even large proteins. This makes several possible applications of this method useful tools for the analysis of the protein structure.

In the last chapter, the aromatic interactions were dealt with in two examples of reactions catalysed by aromatic organic catalysts. It was shown that the interactions between the catalyst and the aromatic portion of the substrate play a significant role in the TS of these reactions. Moreover, in the Quinox-catalysed reaction, the differences in the Gibbs free energy between the two TS structures were calculated which lead to the two different enantiomers. Our theoretical predictions are in very good agreement with the experimental results.

References

[1] S. K. Burley, G. A. Petsko. *Science* **1985**, *229*(4708), 23-28.
[2] S. K. Burley, G. A. Petsko. *Journal of the American Chemical Society* **1986**, *108*(25), 7995-8001.
[3] V. Klusak, Z. Havlas, L. Rulisek, J. Vondrasek, A. Svatos. *Chemistry & Biology* **2003**, *10*(4), 331-340.
[4] V. Spiwok, P. Lipovova, T. Skalova, E. Vondrackova, J. Dohnalek, J. Hasek, B. Kralova. *Journal of Computer-Aided Molecular Design* **2005**, *19*(12), 887-901.
[5] A. Matsushima, T. Fujita, T. Nose, Y. Shimohigashi. *Journal of Biochemistry* **2000**, *128*(2), 225-232.
[6] J. Sponer, P. Jurecka, I. Marchan, F. J. Luque, M. Orozco, P. Hobza. *Chemistry – A European Journal* **2006**, *12*(10), 2854-2865.
[7] J. Rezac, P. Hobza. *Chemistry – A European Journal* **2007**, *13*(10), 2983-2989.
[8] C. Valery, M. Paternostre, B. Robert, T. Gulik-Krzywicki, T. Narayanan, J. C. Dedieu, G. Keller, M. L. Torres, R. Cherif-Cheikh, P. Calvo, F. Artzner. *Proceedings of the National Academy of Sciences of the United States of America* **2003**, *100*(18), 10258-10262.
[9] S. Ito, M. Wehmeier, J. D. Brand, C. Kubel, R. Epsch, J. P. Rabe, K. Mullen. *Chemistry - A European Journal* **2000**, *6*(23), 4327-4342.
[10] E. Huckel. *Zeitschrift fur Physik* **1930**, *60*, 423-456.
[11] W. V. Doering, F. L. Detert. *Journal of the American Chemical Society* **1951**, *73*(2), 876-877.
[12] B. K. Mishra, N. Sathyamurthy. *Journal of Physical Chemistry A* **2007**, *111*(11), 2139-2147.
[13] W. Z. Wang, M. Pitonak, P. Hobza. *Chemphyschem* **2007**, *8*(14), 2107-2111.
[14] E. C. Lee, D. Kim, P. Jurecka, P. Tarakeshwar, P. Hobza, K. S. Kim. *Journal of Physical Chemistry A* **2007**, *111*(18), 3446-3457.
[15] C. A. Hunter, J. K. M. Sanders. *Journal of the American Chemical Society* **1990**, *112*(14), 5525-5534.
[16] M. O. Sinnokrot, C. D. Sherrill. *Journal of Physical Chemistry A* **2003**, *107*(41), 8377-8379.
[17] M. O. Sinnokrot, C. D. Sherrill. *Journal of the American Chemical Society* **2004**, *126*(24), 7690-7697.
[18] K. Pluhackova, P. Jurecka, P. Hobza. *Physical Chemistry Chemical Physics* **2007**, *9*(6), 755-760.
[19] J. Vondrasek, L. Bendova, V. Klusak, P. Hobza. *Journal of the American Chemical Society* **2005**, *127*(8), 2615-2619.
[20] L. Bendova, P. Hobza, J. Vondrasek. *Protein Science* **2008**, *(accepted)*.
[21] L. Bendova, P. Jurecka, P. Hobza, J. Vondrasek. *Journal of Physical Chemistry B* **2007**, *111*(33), 9975-9979.
[22] L. Biedermannova, K. E. Riley, P. Hobza, J. Vondrasek. *Physical Chemistry Chemical Physics* **2008**, *(submitted)*.
[23] A. V. Malkov, A. J. P. S. Liddon, P. Ramirez-Lopez, L. Bendova, D. Haigh, P. Kocovsky. *Angewandte Chemie-International Edition* **2006**, *45*(9), 1432-1435.
[24] A. V. Malkov, P. Ramirez-Lopez, L. Biedermannova, L. Rulisek, L. Dufkova, M. Kotora, F. Zhu , P. Kocovsky. *Journal of the American Chemical Society* **2008**, *(submitted)*.
[25] H. B. Schlegel. *Journal of Computational Chemistry* **1982**, *3*(2), 214-218.
[26] H. Taylor, J. Simons. *Journal of Physical Chemistry* **1985**, *89*(4), 684-688.
[27] J. Baker. *Journal of Computational Chemistry* **1986**, *7*(4), 385-395.
[28] F. Jensen. *Journal of Chemical Physics* **1995**, *102*(17), 6706-6718.
[29] K. E. Riley, P. Hobza. *Journal of Physical Chemistry A* **2007**, *111*(33), 8257-8263.
[30] J. Pittner, P. Hobza. *Chemical Physics Letters* **2004**, *390*(4-6), 496-499.
[31] J. V. Burda, R. Zahradnik, P. Hobza, M. Urban. *Molecular Physics* **1996**, *89*(2), 425-432.
[32] A. J. Misquitta, K. Szalewicz. *Journal of Chemical Physics* **2005**, *122*(21), 214109 (19 pages).

[33] A. Hesselmann, G. Jansen, M. Schutz. *Journal of Chemical Physics* **2005**, *122*(1), 014103 (17 pages).
[34] R. Podeszwa, K. Szalewicz. *Chemical Physics Letters* **2005**, *412*(4-6), 488-493.
[35] J. P. Perdew, K. Burke, M. Ernzerhof. *Physical Review Letters* **1996**, *77*(18), 3865-3868.
[36] A. D. Becke. *Journal of Chemical Physics* **1993**, *98*(7), 5648-5652.
[37] J. M. Tao, J. P. Perdew, V. N. Staroverov, G. E. Scuseria. *Physical Review Letters* **2003**, *91*(14), 146401.
[38] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, M. J. Frisch. *Journal of Physical Chemistry* **1994**, *98*(45), 11623-11627.
[39] K. Burke, M. Ernzerhof, J. P. Perdew. *Chemical Physics Letters* **1997**, *265*(1-2), 115-120.
[40] P. Jurecka, J. Cerny, P. Hobza, D. R. Salahub. *Journal of Computational Chemistry* **2007**, *28*(2), 555-569.
[41] R. Sedlak, P. Jurecka, P. Hobza. *Journal of Chemical Physics* **2007**, *127*(7), 075104 (3 pages).
[42] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, P. A. Kollman. *Journal of the American Chemical Society* **1995**, *117*(19), 5179-5197.
[43] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, M. Karplus. *Journal of Computational Chemistry* **1983**, *4*(2), 187-217.
[44] E. Lindahl, B. Hess, D. van der Spoel. *Journal of Molecular Modeling* **2001**, *7*(8), 306-317.
[45] D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, H. J. C. Berendsen. *Journal of Computational Chemistry* **2005**, *26*(16), 1701-1718.
[46] A. Vibok, I. Mayer. *International Journal of Quantum Chemistry* **1992**, *43*(6), 801-811.
[47] I. Mayer. *International Journal of Quantum Chemistry* **1998**, *70*(1), 41-63.
[48] S. F. Boys, F. Bernardi. *Molecular Physics* **1970**, *19*(4), 553-566.
[49] D. G. Truhlar. *Chemical Physics Letters* **1998**, *294*(1-3), 45-48.
[50] P. L. Fast, M. L. Sanchez, D. G. Truhlar. *Journal of Chemical Physics* **1999**, *111*(7), 2921-2926.
[51] A. Halkier, T. Helgaker, P. Jorgensen, W. Klopper, H. Koch, J. Olsen, A. K. Wilson. *Chemical Physics Letters* **1998**, *286*(3-4), 243-252.
[52] W. Klopper. *Journal of Chemical Physics* **2004**, *120*(23), 10890-10895.
[53] P. Jurecka, P. Nachtigall, P. Hobza. *Physical Chemistry Chemical Physics* **2001**, *3*(20), 4578-4582.
[54] P. E. Smith, B. M. Pettitt. *Journal of Physical Chemistry* **1994**, *98*(39), 9700-9711.
[55] C. J. Cramer, D. G. Truhlar. *Science* **1992**, *256*(5054), 213-217.
[56] M. Born. *Zeitschrift für Physik* **1920**, *1*, 45-48.
[57] L. Onsager. *Journal of the American Chemical Society* **1936**, *58*, 1486-1493.
[58] O. Tapia, O. Goscinski. *Molecular Physics* **1975**, *29*(6), 1653-1661.
[59] J. B. Foresman, T. A. Keith, K. B. Wiberg, J. Snoonian, M. J. Frisch. *Journal of Physical Chemistry* **1996**, *100*(40), 16098-16104.
[60] S. Miert, E. Scrocco, J. Tomasi. *Chemical Physics* **1981**, *55*(1), 117-129.
[61] A. Klamt, G. Schuurmann. *Journal of the Chemical Society-Perkin Transactions 2* **1993**, (5), 799-805.
[62] A. Schafer, A. Klamt, D. Sattel, J. C. W. Lohrenz, F. Eckert. *Physical Chemistry Chemical Physics* **2000**, *2*(10), 2187-2193.
[63] R. Constanciel, R. Contreras. *Theoretica Chimica Acta* **1984**, *65*(1), 1-11.
[64] D. Bashford, D. A. Case. *Annual Review of Physical Chemistry* **2000**, *51*, 129-152.
[65] C. J. Cramer, D. G. Truhlar. *Journal of Computer-Aided Molecular Design* **1992**, *6*(6), 629-666.
[66] C. C. Chambers, G. D. Hawkins, C. J. Cramer, D. G. Truhlar. *Journal of Physical Chemistry* **1996**, *100*(40), 16385-16398.
[67] C. D. Tatko, M. L. Waters. *Journal of the American Chemical Society* **2002**, *124*(32), 9372-9373.
[68] C. B. Anfinsen. *Science* **1973**, *181*(4096), 223-230.
[69] D. Baker. *Nature Structural Biology* **1998**, *5*(12), 1021-1024.

[70] R.H.Garrett, CM. Grisham, Biochemistry (with BiochemistryNow and InfoTrac), 3rd Edition, Brooks/Cole, 2005.
[71] I. Y. Torshin, I. T. Weber, R. W. Harrison. *Protein Engineering* **2002**, *15*(5), 359-363.
[72] C. Brandon, J. Tooze. *Introduction to Protein Structure,* Garland Publishing, New York/London, **1991**.
[73] K. A. Dill. *Biochemistry* **1990**, *29*(31), 7133-7155.
[74] J. Tsai, M. Gerstein, M. Levitt. *Protein Science* **1997**, *6*(12), 2606-2616.
[75] E. A. Meyer, R. K. Castellano, F. Diederich. *Angewandte Chemie-International Edition* **2003**, *42*(11), 1210-1250.
[76] C. Tanford. *Science* **1978**, *200*(4345), 1012-1018.
[77] G. D. Rose, A. R. Geselowitz, G. J. Lesser, R. H. Lee, M. H. Zehfus. *Science* **1985**, *229*(4716), 834-838.
[78] K. Muller-Dethlefs, P. Hobza. *Chemical Reviews* **2000**, *100*(1), 143-167.
[79] P. Hobza, J. Sponer. *Journal of the American Chemical Society* **2002**, *124*(39), 11802-11808.
[80] P. Hobza, R. Zahradnik, K. Muller-Dethlefs. *Collection of Czechoslovak Chemical Communications* **2006**, *71*(4), 443-531.
[81] S. K. Burley, G. A. Petsko. *Science* **1985**, *229*(4708), 23-28.
[82] S. Tsuzuki, T. Uchimaru, K. Sugawara, M. Mikami. *Journal of Chemical Physics* **2002**, *117*(24), 11216-11221.
[83] M. O. Sinnokrot, C. D. Sherrill. *Journal of Physical Chemistry A* **2006**, *110*(37), 10656-10668.
[84] M. Vendruscolo, E. Paci, C. M. Dobson, M. Karplus. *Nature* **2001**, *409*(6820), 641-645.
[85] K. Lindorff-Larsen, M. Vendruscolo, E. Paci, C. M. Dobson. *Nature Structural & Molecular Biology* **2004**, *11*(5), 443-449.
[86] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne. *Nucleic Acids Research* **2000**, *28*(1), 235-242.
[87] G. Neshich, R. C. Togawa, A. L. Mancini, P. R. Kuser, M. E. B. Yamagishi, G. Pappas, W. V. Torres, T. F. E. Campos, L. L. Ferreira, F. M. Luna, A. G. Oliveira, R. T. Miura, M. K. Inoue, L. G. Horita, D. F. de Souza, F. Dominiquini, A. Alvaro, C. S. Lima, F. O. Ogawa, G. B. Gomes, J. F. Palandrani, G. F. dos Santos, E. M. de Freitas, A. R. Mattiuz, I. C. Costa, C. L. de Almeida, S. Souza, C. Baudet, R. H. Higa. *Nucleic Acids Research* **2003**, *31*(13), 3386-3392.
[88] A. Halkier, T. Helgaker, P. Jorgensen, W. Klopper, H. Koch, J. Olsen, A. K. Wilson. *Chemical Physics Letters* **1998**, *286*(3-4), 243-252.
[89] K. E. Riley, K. M. Merz. *Journal of Physical Chemistry B* **2006**, *110*(32), 15650-15653.
[90] J. L. Gao, T. R. Furlani. *IEEE Computational Science & Engineering* **1995**, *2*(3), 24-33.
[91] J. Vondrasek, T. Kubar, F. E. Jenney, M. W. W. Adams, M. Kozisek, J. Cerny, V. Sklenar, P. Hobza. *Chemistry-a European Journal* **2007**, *13*(32), 9022-9027.
[92] M. Vendruscolo, E. Paci, C. M. Dobson, M. Karplus. *Nature* **2001**, *409*(6820), 641-645.
[93] K. Lindorff-Larsen, M. Vendruscolo, E. Paci, C. M. Dobson. *Nature Structural & Molecular Biology* **2004**, *11*(5), 443-449.
[94] C. Berezin, F. Glaser, J. Rosenberg, I. Paz, T. Pupko, P. Fariselli, R. Casadio, N. Ben-Tal. *Bioinformatics* **2004**, *20*(8), 1322-1324.
[95] C. J. Chen, L. Li, Y. Xiao. *Physical Review E* **2006**, *73*(4), 041926 (7 pages).
[96] T. Haliloglu, I. Bahar, B. Erman. *Physical Review Letters* **1997**, *79*(16), 3090-3093.
[97] I. Bahar, A. R. Atilgan, B. Erman. *Folding & Design* **1997**, *2*(3), 173-181.
[98] P. Jurecka, J. Cerny, P. Hobza, D. R. Salahub. *Journal of Computational Chemistry* **2007**, *28*(2), 555-569.
[99] P. K. Weiner, P. A. Kollman. *Journal of Computational Chemistry* **1981**, *2*(3), 287-303.
[100] M. Isobe, H. Shimizu, Y. Hiwatari. *Journal of the Physical Society of Japan* **2001**, *70*(5), 1233-1236.
[101] J. L. Gao, T. R. Furlani. *IEEE Computational Science & Engineering* **1995**, *2*(3), 24-33.
[102] A. Schafer, A. Klamt, D. Sattel, J. C. W. Lohrenz, F. Eckert. *Physical Chemistry Chemical Physics* **2000**, *2*(10), 2187-2193.

[103] A. Schafer, A. Klamt, D. Sattel, J. C. W. Lohrenz, F. Eckert. *Physical Chemistry Chemical Physics* **2000**, *2*(10), 2187-2193.
[104] R. Constanciel, R. Contreras. *Theoretica Chimica Acta* **1984**, *65*(1), 1-11.
[105] J. W. Neidigh, R. M. Fesinmeyer, N. H. Andersen. *Nature Structural Biology* **2002**, *9*(6), 425-430.
[106] M. Lund, B. Jonsson, C. E. Woodward *Journal of Chemical Physics* **2007**, *126*, 225103.
[107] J. Vondrasek, L. Bendova, V. Klusak, P. Hobza. *Journal of the American Chemical Society* **2005**, *127*(8), 2615-2619.
[108] W. Z. Wang, M. Pitonak, P. Hobza. *Chemphyschem* **2007**, *8*(14), 2107-2111.
[109] E. C. Lee, D. Kim, P. Jurecka, P. Tarakeshwar, P. Hobza, K. S. Kim. *Journal of Physical Chemistry A* **2007**, *111*(18), 3446-3457.
[110] G. L. Duan, V. H. Smith, D. F. Weaver. *Chemical Physics Letters* **1999**, *310*(3-4), 323-332.
[111] J. G. Cheng, C. M. Kang, W. L. Zhu, X. M. Luo, C. M. Puah, K. X. Chen, J. H. Shen, H. L. Jiang. *Journal of Organic Chemistry* **2003**, *68*(19), 7490-7495.
[112] J. B. O. Mitchell, C. L. Nandi, I. K. Mcdonald, J. M. Thornton, S. L. Price. *Journal of Molecular Biology* **1994**, *239*(2), 315-331.
[113] J. Singh, J. M. Thornton. *Journal of Molecular Biology* **1990**, *211*(3), 595-615.
[114] K. Watanabe, T. Masuda, H. Ohashi, H. Mihara, Y. Suzuki. *European Journal of Biochemistry* **1994**, *226*(2), 277-283.
[115] Y. Suzuki, K. Oishi, H. Nakano, T. Nagayama. *Applied Microbiology and Biotechnology* **1987**, *26*(6), 546-551.
[116] Y. Suzuki. *Proceedings of the Japan Academy Series B-Physical and Biological Sciences* **1999**, *75*(6), 133-137.
[117] H. N. Cheng, F. A. Bovey. *Biopolymers* **1977**, *16*(7), 1465-1472.
[118] C. Grathwohl, K. Wuthrich. *Biopolymers* **1981**, *20*(12), 2623-2633.
[119] K. E. Riley, G. L. Cui, K. M. Merz. *Journal of Physical Chemistry B* **2007**, *111*(20), 5700-5707.
[120] A. J. Misquitta, K. Szalewicz. *Journal of Chemical Physics* **2005**, *122*(21), 214109 (19 pages).
[121] A. Hesselmann, G. Jansen, M. Schutz. *Journal of Chemical Physics* **2005**, *122*(1).
[122] R. A. Laskowski, N. Luscombe. *Atlas of Protein Side-Chain Interactions,* http://www.biochem.ucl.ac.uk/bsm/sidechains/ **2007**.
[123] J. Singh, J. M. Thornton. *Atlas of Protein Side-Chain Interactions, Vols. I & II,* IRL press, Oxford, **1992**.
[124] P. Jurecka, P. Hobza. *Chemical Physics Letters* **2002**, *365*(1-2), 89-94.
[125] A. Hesselmann, G. Jansen, M. Schutz. *Journal of the American Chemical Society* **2006**, *128*(36), 11730-11731.
[126] A. Hesselmann, G. Jansen, M. Schutz. *Journal of Chemical Physics* **2005**, *122*(1), 014103 (17 pages).
[127] T. Steiner, G. Koellner. *Journal of Molecular Biology* **2001**, *305*(3), 535-557.
[128] A. Hesselmann, G. Jansen. *Chemical Physics Letters* **2002**, *357*(5-6), 464-470.
[129] G. Toth, C. R. Watts, R. F. Murphy, S. Lovas. *Proteins-Structure Function and Genetics* **2001**, *43*(4), 373-381.
[130] A. V. Malkov, M. Orsini, D. Pernazza, K. W. Muir, V. Langer, P. Meghani, P. Kocovsky. *Organic Letters* **2002**, *4*(6), 1047-1049.
[131] A. V. Malkov, M. Bell, M. Orsini, D. Pernazza, A. Massa, P. Herrmann, P. Meghani, P. Kocovsky. *Journal of Organic Chemistry* **2003**, *68*(25), 9659-9668.
[132] A. V. Malkov, M. Bell, M. Vassieu, V. Bugatti, P. Kocovsky. *Journal of Molecular Catalysis A: Chemical* **2003**, *196*(1-2), 179-186.
[133] A. V. Malkov, L. Dufkova, L. Farrugia, P. Kocovsky. *Angewandte Chemie-International Edition* **2003**, *42*(31), 3674-3677.
[134] A. V. Malkov, M. Bell, F. Castelluzzo, P. Kocovsky. *Organic Letters* **2005**, *7*(15), 3219-3222.
[135] S. E. Denmark, J. P. Fu, D. M. Coe, X. P. Su, N. E. Pratt, B. D. Griedel. *Journal of Organic Chemistry* **2006**, *71*(4), 1513-1522.

[136] E. C. Lee, D. Kim, P. Jurecka, P. Tarakeshwar, P. Hobza, K. S. Kim. *Journal of Physical Chemistry A* **2007**, *111*(18), 3446-3457.
[137] M. O. Sinnokrot, C. D. Sherrill. *Journal of the American Chemical Society* **2004**, *126*(24), 7690-7697.
[138] F. Jensen. *Introduction to Computational Chemistry,* John Wiley & Sons, **1999**.
[139] T. L. Hill. *An introduction to statistical thermodynamics.,* Dover Publications Inc., New York, **1960**.
[140] R. Ahlrichs, M. Bar, M. Haser, H. Horn, C. Kolmel. *Chemical Physics Letters* **1989**, *162*(3), 165-169.
[141] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W.Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople. Gaussian 03, Revision C.02. 2004.
[142] P. Jurecka, J. Cerny, P. Hobza, D. R. Salahub. *Journal of Computational Chemistry* **2007**, *28*(2), 555-569.
[143] MOLPRO, a package of ab initio programs designed by H.-J. Werner and P. J. Knowles, version 2002.1, R. D. Amos, A. Bernhardsson, A. Berning, P. Celani, D. L. Cooper, M. J. O. Deegan, A. J. Dobbyn, F. Eckert, C. Hampel, G. Hetzer, P. J. Knowles, T. Korona, R. Lindh, A. W. Lloyd, S. J. McNicholas, F. R. Manby, W. Meyer, M. E. Mura, A. Nicklass, P. Palmieri, R. Pitzer, G. Rauhut, M. Schutz, U. Schumann, H. Stoll, A. J. Stone, R. Tarroni, T. Thorsteinsson, and H.-J. Werner.
[144] P. K. Weiner, P. A. Kollman. *Journal of Computational Chemistry* **1981**, *2*(3), 287-303.

List of Abbreviations

AA – Amino Acid

AO – Atomic Orbital

BSSE – Basis Set Superposition Error

CBS – Complete Basis Set

CCSD(T) – Coupled Clusters with Single, Double and Perturbative Triple Excitations

CCSDT – Coupled Clusters with Single, Double and Triple Excitations

COSMO – Continuum Solvation Model

CPU – Central Processor Unit

DFT – Density Functional Theory

DFT-D – Density Functional Theory with Empirical Dispersion Correction

DNA – Deoxyribonucleic Acid

FCA – Frozen Core Approximation

FMA – Formamide

GB – Generalised Born Solvent Model

GB/SA – Generalised Born Solvent Model with Cavitation and van der Waals Corrections Based on the Surface–Accessible Area Model

GGA – Generalised Gradient Approximation

H-bond – Hydregen Bond

HF – Hartree-Fock Method

CHA – Chemical Hamiltonian Approach

IEM – Interaction Energy Matrix

LDA – Local Density Approximation

MD – Molecular Dynamics

MM – Molecular Mechanics

MO – Molecular Orbital

MP – Møller-Plesset Method

NMA – *N*-methylacetamide

NMF – *N*-methylformamide

NMR – Nuclear Magnetic Resonance

PCM – Polarisable Continuum Model

PDB – Protein Databank

PES – Potential Energy Surface

QM – Quantum Mechanics

RI – Resolution of Identity Approximation

SAPT – Symmetry Adapted Perturbation Theory

SAPT(DFT) – Symmetry Adapted Perturbation Theory Based on Kohn-Sham Orbitals

SCC-DFTB – Self-Consistent Charge Density Functional Tight-Binding Method

SCC-DFT-D – Self-Consistent Charge Density Functional Tight-Binding Method Augmented with an Empirical Dispersion Term

SCRF – Self-Consistent Reaction Field Method

TS – Transition State

TSE – Transition State Ensemble

vdW – van der Waals

Molecular Modelling Packages Used

The DFT calculations were performed using the Turbomole program package (versions 6-9)[140] and Gaussian03.[141] For DFT-D calculations, a computer code developed in our own department was used.[142] The MP2 calculations were carried out in the Turbomole while the CCSD(T) calculations were performed using the MOLPRO program.[143] Empirical force field calculations were done in the Amber package;[144] the SAPT(DFT) calculations were performed in the Gaussian03 and MOLPRO programs.

Appendix

A. J. Vondrášek, L. Bendová, V. Klusák, P. Hobza: Unexpectedly strong energy stabilisation inside the hydrophobic core of small protein rubredoxin mediated by aromatic residues: Correlated *ab initio* quantum chemical calculations. *Journal of the American Chemical Society* **2005**, *127*(8), 2615-2619.

B. L. Bendová, P. Hobza, J. Vondrášek: Identifying stabilizing key residues in proteins using interresidue interaction energy matrix. *Protein Science* **2008**, *(accepted).*

C. L. Bendová, P. Jurečka, P. Hobza, J. Vondrášek: Model of Peptide Bond–Aromatic Ring Interaction: Correlated Ab Initio Quantum Chemical Study. *Journal of Physical Chemistry B* **2007** *111*(33), 9975-9979.

D. L. Biedermannová, K. E. Riley, P. Hobza, J. Vondrášek: Another role of Proline: Stabilization interactions in proteins concerning proline and residues of aromatic characters. *Physical Chemistry Chemical Physics* **2008**, *(submitted).*

E. A. V. Malkov, A. J. P. S. Liddon, P. Ramírez-López, L. Bendová, D. Haigh, P. Kočovský: Remote chiral induction in the organocatalytic hydrosilylation of aromatic ketones and ketimines. *Angewandte Chemie – International Edition* **2006**, *45*(9), 1432-1435.

F. A. V. Malkov, P. Ramírez-López, L. Biedermannová, L. Rulíšek, L. Dufková, M. Kotora, F. Zhu, Kočovský P: On the Mechanism of Asymmetric Allylation of Aldehydes with Allyltrichlorosilanes Catalyzed by Quinox, a Chiral Isoquinoline N-Oxide. *Journal of the American Chemical Society* **2008**, *(submitted).*

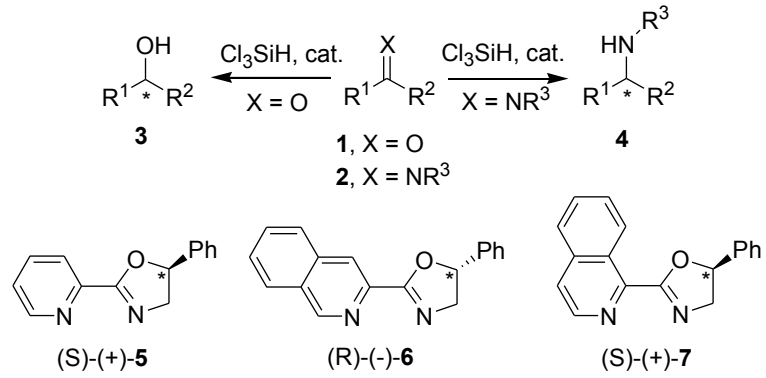
6 Aromatic Interactions in Organocatalysis

6.1 Introduction

The preceding chapters dealt with aromatic interactions in biological molecules, namely in proteins, and showed that these interactions play a significant role in the stabilisation of the tertiary structure of these biologically important molecules. In this chapter, the role which these interactions can play in the stabilisation of complexes of synthetic origin will be investigated in greater detail, specifically of the transition-state (TS) complexes in reactions catalysed by asymmetric organic catalysts containing an aromatic ring.

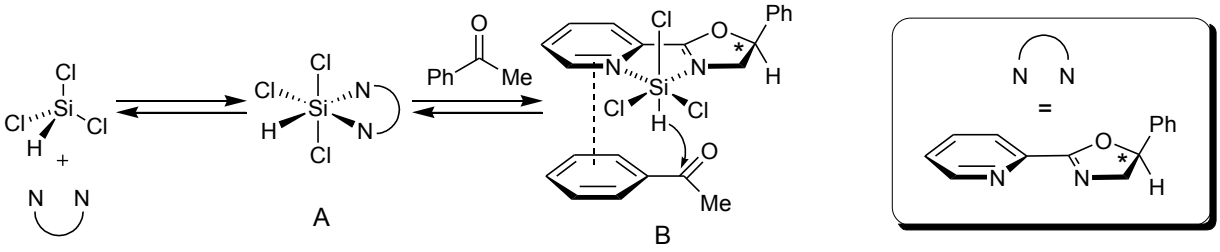
This project has originated in the group of Prof Kočovský, which focuses on the activation of organosilicon reagents by Lewis bases. Most of these Lewis-basic catalysts, like the reaction substrates, are aromatic; it has been proposed that aromatic–aromatic interactions between the substrate and the catalyst play an important role in the TS of the reaction. Therefore, the decision was taken to study two different reactions – the asymmetric reduction of ketones and ketimines and the asymmetric allylation of aldehydes – catalysed by two different catalysts. A separate section of this chapter is devoted to each of these reactions.

The first of the two reactions is the hydrosilylation of ketones **1** and ketimines **2**, yielding chiral alcohols **3** and secondary amines **4**, catalysed by pyridyloxazolines **5**, **6** and **7**, respectively (see Scheme 6.1). These catalysts were shown to yield very good enantiomeric excess (70-80%) with a series of aromatic ketones. On the other hand, the reduction of non-aromatic ketones resulted in the formation of a racemic product.



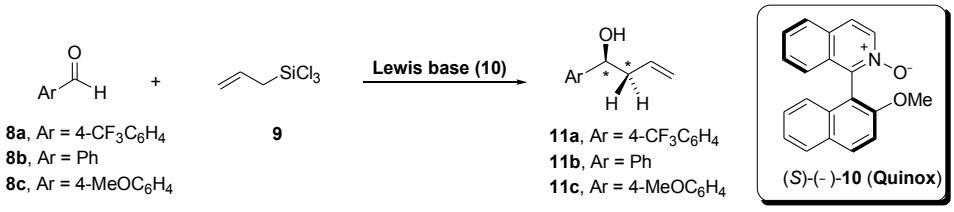
Scheme 6.1: The catalytic reduction of ketones **1** and ketimines **2**.

Based on these as well as other experimental observations (for details, see Appendix E), the following mechanistic picture of the reaction was proposed (see Scheme 6.2). Upon the *N,N*-chelation by an oxazoline catalyst, the Cl_3SiH molecule creates an activated hydrosilylating species (chelate A in Scheme 6.2). The substrate aromatic ketone then approaches the catalyst–Cl_3SiH complex from the less hindered side, *i.e.* the one without the phenyl substituent, forming a TS structure (complex B in Scheme 6.2). This putative TS is stabilised by aromatic interactions between the pyridine ring of the catalyst and the aromatic substrate. To support this hypothesis, a scan of the interaction energy was performed for a model system of the substrate–catalyst interaction and the optimum interaction geometry and corresponding interaction energy identified.



Scheme 6.2: The mechanism of hydrosilylation.

The other reaction studied by computational methods in this chapter is the asymmetric allylation of aldehydes **8** with allyltrichlorosilane **9**, catalysed by Quinox **10**, a chiral isoquinoline-*N*-oxide, yielding enantiomerically enriched homoallylic alcohols **11** (see Scheme 6.3). Similarly to the previous reaction, this reaction also proceeds for aromatic substrates only. An interesting feature of Quinox is that, unlike the majority of the catalysts reported to date, it exhibits a noteworthy preference towards electron-poor aldehydes (96% *ee* for aldehyde **8a**), whereas for electron-rich substrates it yields much lower *ee* and the reaction is also much slower.



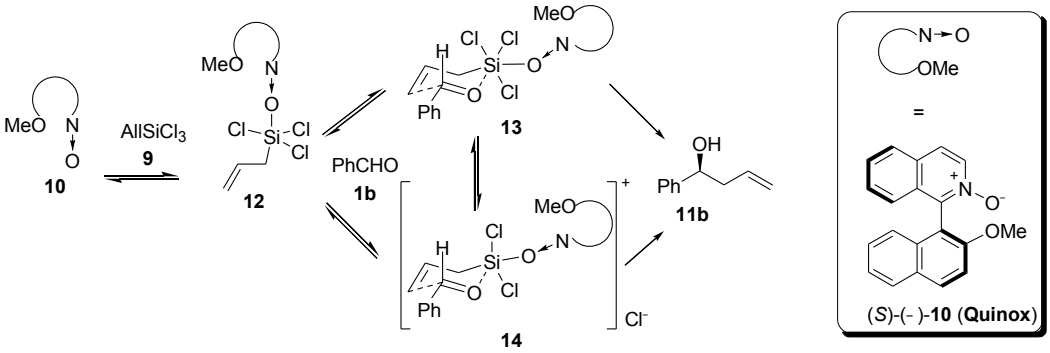
8a, Ar = 4-$CF_3C_6H_4$
8b, Ar = Ph
8c, Ar = 4-$MeOC_6H_4$

9

11a, Ar = 4-$CF_3C_6H_4$
11b, Ar = Ph
11c, Ar = 4-$MeOC_6H_4$

Scheme 6.3: The allylation of aldehydes catalysed by chiral Lewis bases.

The mechanistic scenario, suggested on the basis of kinetic experiments, is in this case slightly more complicated (for details of these experiments, see Appendix F). As depicted in Scheme 6.4, allyltrichlorosilane **9** and Quinox **10** reversibly form complex **12**. The next reaction step, the binding of the aldehyde, can proceed via an associative (TS **13**) or dissociative (TS **14**) mechanism. Although some indications favour the neutral TS **13,** the kinetic data alone do not provide a definite answer. In order to discriminate these two possible mechanisms from each other, the TS barriers were computationally determined for both channels (leading to (*R*)- and (*S*)- products) in both the associative and the dissociative mechanism.



Scheme 6.4: The allylation of benzaldehyde, catalysed by Quinox **10**.

6.2 Asymmetric Reduction Using Oxazoline Catalysts

In order to assess the potential role of the aromatic interactions in shaping the TS of the reduction reaction, a computational analysis of a pyridine–acetophenone model complex was carried out (Figure 6.1). In this model complex, a series of successive scans were performed, according to the steering angle θ, the distance between the planes r_1, and the parallel displacement of the two molecules r. Starting from a strictly parallel structure of the complex, with the steering angle $\theta = 0°$ (the pyridine's N atom and the acetyl group aligned) and the interplanar distance $r_1 = 3.8$ Å, each of the successive scans began from the geometry of the interaction-energy (E_{int}) minimum of the previous scan. The geometries of the monomers

were optimised by the MP2/cc-pVTZ method and kept fixed throughout the scans. The interaction energy at each point was computed using the MP2/aug-cc-pVDZ and CP-corrected for the BSSE.

The lowest interaction energy found by this scheme was E_{int} = -6.3 kcal/mol, and the geometry has been depicted in Figure 6.1. The increased stabilisation when compared to *e.g.* the benzene dimer, where the interaction energy at the CCSD(T)/CBS level amounts to -2.73 kcal/mol, can be attributed to several factors. It has been shown that the substitution of either electron-donating or electron-withdrawing groups leads to a more favourable interaction energy due to the increased polarisability of the monomers. This effect is probably also responsible for the large interaction energy observed in our model system. Other contributions include the effect of the favourable interaction between the CO group of acetophenone and the pyridine ring as well as the interaction between the pyridine's N atom π-orbital and the partial positive charge of the acetophenone methyl group. The considerably large interaction energy in the model system supports the assumed important role of the π–π stacking contribution in the TS of the hydrosilylation reaction. Moreover, the most stable geometry found for the model system is in good agreement with the structure of the putative TS.

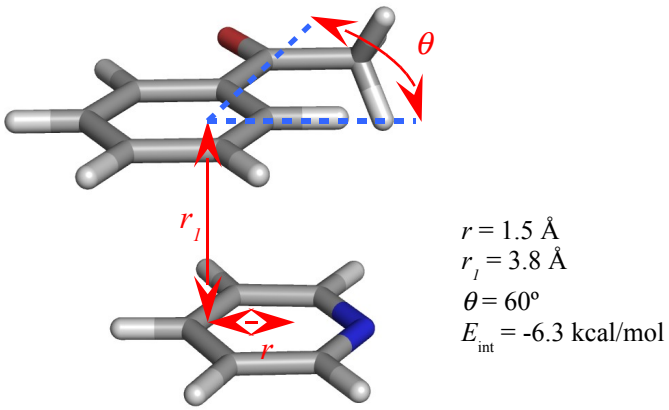


Figure 6.1: The arene–arene interactions of acetophenone and pyridine in the model of the TS.

6.3 Asymmetric Allylation Using Quinox

In order to capture the small differences between the two possible mechanisms and their (*R*)- and (*S*)- channels,* it is necessary to describe the structure of the TS as accurately as possible. So as to describe the dispersion energy correctly in these relatively large systems, the DFT-D method was used (for details, see Appendix F). To account for the solvation effects, the COSMO model of the solvent was used with its dielectric constant corresponding to acetonitrile ($\varepsilon_r = 36.6$). The geometry optimisations of the reactants as well as of the TS were performed using the DFT-D/PBE/6-31G(d). The TS search was done using the recommended protocol described in Chapter 2, the interpolated reaction coordinate was defined by the breaking of the Si–C bond and the forming of the C–C bond (see Scheme 6.4).

Due to the lack of analytic second derivatives in the COSMO method, an approximate method was used for obtaining the TS in solution. First, the TS structure was optimised according to the procedure described above in the gas phase. To account for the solvation effects, the length of the Si–C and C–C bonds which define the reaction coordinate was constrained and all other degrees of freedom optimised using the DFT-D/COSMO method.

At the final optimised geometry, the Gibbs free energy was calculated as the sum of these contributions:

$$\Delta G = \Delta E + \Delta G_{\text{solv}} + (\Delta E_{\text{ZPE}} - RT \ln(q_{\text{trans}}\, q_{\text{rot}}\, q_{\text{vib}})), \tag{6.1}$$

where ΔE is the gas-phase energy of the system at the DFT-D/PBE/TZVPP level, ΔG_{solv} is the solvation free energy at the DFT-D/PBE/6-31G* level (calculated as the difference in the energy of the system in the gas phase and in the COSMO solvation model) and the (ΔE_{ZPE} - RT ln(q_{trans} q_{rot} q_{vib})) term is the zero-point energy, thermal corrections to the Gibbs free energy and entropic term obtained from a frequency calculation by means of the DFT-D/PBE/6-31G* method at 298 K temperature and 1 atm pressure, using an ideal gas approximation. The free energy calculated with this equation can be considered a good approximation to the free energy in a diluted solution.

* According to the equation $\Delta G = -RT\ lnK$, an order difference of the reaction rates corresponds to a 1.36 kcal/mol difference in the ΔG of the TS of the two reactions.

Approximately twenty structural variants were initially examined for each of the reaction mechanisms using (*R*)-(+)-Quinox **10** and benzaldehyde **8b** as a model substrate to find the most stable TS structures. For these structures, the corresponding reactant and products were ascertained and the ΔG was calculated as described above. The structures found for the associative mechanism are shown in Figure 6.2 and the results for both associative and dissociative pathways are summarised in Table 6.1.

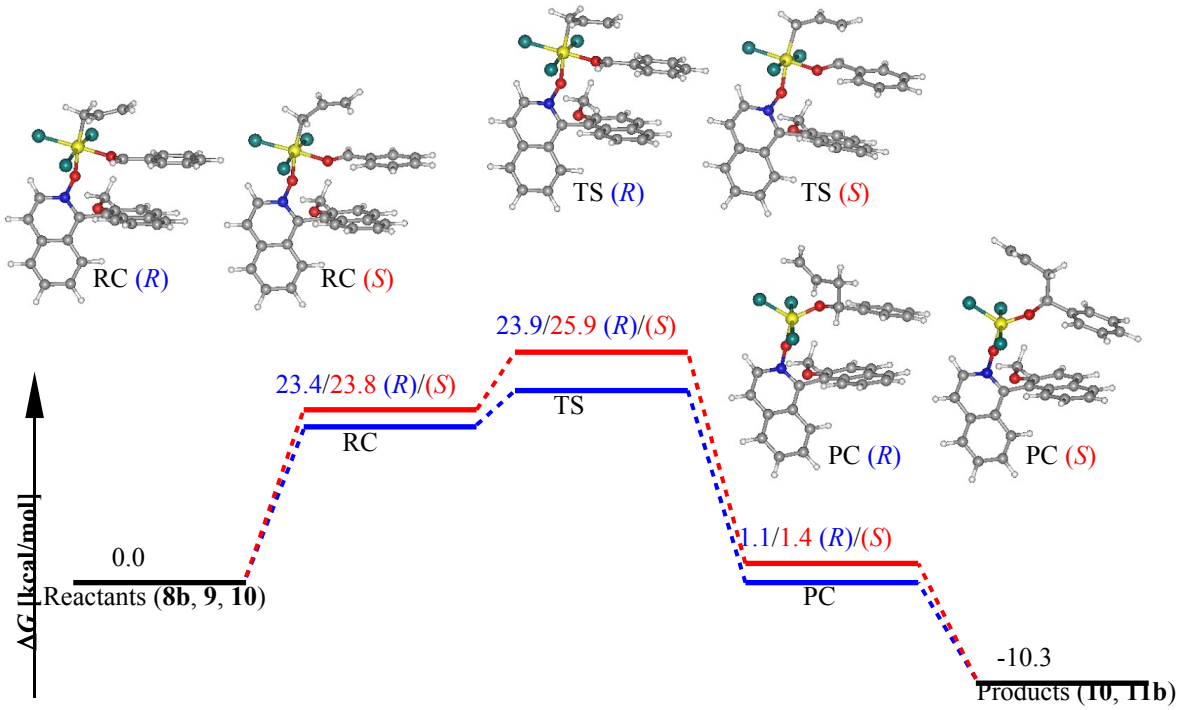


Figure 6.2. The reaction coordinate for the associative pathway of allylation of benzaldehyde **8b** using allyltrichlorosilane **9** catalysed by Quinox **10**, forming alcohol **11b**. The geometries of the most stable reactant complexes (RC), transition states (TS), and product complexes (PC) of both (*R*)- and (*S*)- product-leading channels. The ΔG values are in kcal/mol; the graph is not proportional.

Table 6.1: The calculated thermochemical data for the reaction of benzaldehyde **8b** with allyltrichlorosilane **9**, catalysed by Quinox **10**. All values are in kcal/mol.

Mechanism	Product configuration	ΔG, RC	$\Delta G^{\neq}$, TS	ΔG, PC
Associative	*R*	23.4	23.9	1.1
Associative	*S*	23.8	25.9	1.4
Dissociative	*R*	19.3	26.0	1.1
Dissociative	*S*	19.8	24.5	1.4

According to this analysis, the reaction commences with the formation of a transient reactant complex (RC). For an associative mechanism, this complex is higher in energy than the assembly of the isolated reactants and the catalyst by 23.4/23.8 kcal/mol (*R*/*S*) (Table 6.1, Rows 2 and 3). For a dissociative mechanism, the formation of the RC is slightly less energetically demanding, with values of 19.3/19.8 kcal/mol (*R*/*S*) (Table 6.1, Rows 4 and 5). The calculated TS barriers are: $\Delta G^{\neq}_{assoc}$ = 23.9/25.9 kcal/mol (*R*/*S*) for the associative mechanism and $\Delta G^{\neq}_{dissoc}$ = 26.0/24.5 kcal/mol (*R*/*S*) for the dissociative mechanism. The overall reaction thermodynamics, after the catalyst has dissociated from the product complex (PC), is ΔG = -10.3 kcal/mol.

The associative mechanism predicts the formation of (*R*)-**11b** in 97% *ee*, which is in good agreement with the experimental value of 87% *ee*. Although the dissociative route favours the formation of the opposite, *i.e*. (*S*)-enantiomer, this mechanism has an uncertainty in the translational entropy of the Cl^- ion, which needs to be considered. The free energy of chloride solvation in acetonitrile, which has been estimated as $\Delta G_{solv}(Cl^-)$ = -84.8 kcal/mol, was calculated in the same way as for the rest of the systems, except for the fact that the Sakur-Tetrode equation was used for the calculation of the translational entropy. The barriers for both mechanisms are close enough to make them both plausible reaction pathways. Assuming that both mechanisms operated concurrently, the calculations would predict the formation of (*R*)-**11b** in 53% *ee*.

The difference between the TS barriers in the associative mechanism, the $\Delta G^{\neq}_{assoc,}$ between the (*R*)- and (*S*)-enantiomers, which amounts to 2.0 kcal/mol, can be decomposed into the differences between several terms as follows (in kcal/mol): gas-phase DFT energies ΔE_{gp} = +0.3; solvation free energies ΔG_{solv} = -0.7; zero-point energy corrections and gas-phase entropic terms -0.5 kcal/mol; and dispersion contribution (the dispersion correction part of the DFT-D interaction energy) ΔE_{disp} = -1.1 kcal/mol. As the last term can be attributed mainly to the aromatic interactions between the catalyst and the substrate, it can be concluded that the

enantiodifferentiation in the QUINOX-catalysed reaction lies in the main in the more favourable aromatic interactions in the (*R*)-enantiomer leading TS complex.

Like in the hydrogenation reaction, the calculations showed that the aromatic groups of the catalyst and of the substrate are arranged in a stacked orientation in the TS. It is the effect of the Quinox catalyst, including the favourable aromatic interactions, which makes the TS of the catalysed reaction lower in energy by 3-5 kcal/mol as compared to the reaction catalysed only by simple pyridine *N*-oxide (where $\Delta G^{\#} = 27.5$ kcal/mol). This value is comparable with the strength of the aromatic interaction between benzaldehyde and the catalyst expected in the TS arrangement.

The described strategy was further employed to evaluate the TS barriers in a reaction with a more electron-rich substrate, the *p*-methoxybenzaldehyde **8c**, which exhibits lower reactivity as well as lower *ee* (12% *ee* observed experimentally) in the Quinox-catalysed allylation. In this case, the associative mechanism, where the TS barriers are $\Delta G^{\neq}_{\text{assoc}} = 26.1/26.9$ kcal/mol for the (*R*)- and (*S*)- channels, is clearly more favoured over the dissociative pathway ($\Delta G^{\neq}_{\text{dissoc}} = 29.3/29.7$ kcal/mol for (*R*)-/(*S*)- isomers). The higher TS barrier of the reaction of *p*-methoxybenzaldehyde **8c** as compared to the unsubstituted benzaldehyde **8b** and the narrower energy gap between the (*R*)- and (*S*)- reaction channels in the associative mechanism (only 0.8 kcal/mol, which corresponds to 62% *ee*) correlate well with the experimentally observed slower reaction and lower enantioselectivity (45% *ee* observed experimentally at 0 °C).

6.4 Conclusions

In this chapter, we have shown that aromatic interactions play a significant role in the two studied reactions catalysed by chiral organic catalysts.

In the reduction of aromatic ketones and ketimines catalysed by pyridyloxazoline catalysts, the putative arene–arene stacking interactions between the pyridyl group of the catalyst and the substrate have been investigated directly by scans of the interaction energy. The interaction energy in the pyridine–acetophenone model system has been shown to amount to as high as -6.3 kcal/mol at the MP2/aug-cc-pVDZ level. The computationally identified optimum geometry of the interaction is in agreement with the putative TS structure, which, along with the relatively high strength of the interaction, provides further evidence to the proposed reaction mechanism.

In the asymmetric allylation of aldehydes catalysed by Quinox, we have focused on the discrimination between two possible mechanisms, associative and dissociative, and the origin of the enantiodifferentiation. The calculations have provided plausible structures of transition states as well as free energy profiles for all reaction channels. We have shown that the reaction is likely to proceed via an associative pathway in a cyclic transition state. The experimentally observed preference for the (*R*)- alcohol has been reproduced, and the calculated 97% *ee* is in good agreement with the experimental value of 87% *ee*. The calculations also reproduce the decrease of enantioselectivity in reaction with electron-rich substrates, *i.e.* the narrowing of the energy gap between the (*R*)- and (*S*)- reaction channels in the associative mechanism.

The excellent agreement between the experimental and theoretical data made it possible to carry out the decomposition of molecular energy into various contributions (solvation, entropic, and dispersion energies) and estimate their role in the enantioselectivity of the reaction. The attractive aromatic interactions between the catalyst and the substrate aldehyde have been shown to play a major role in the enantiodifferentiation.

5 A Detailed Investigation of Selected Interactions

5.1 Introduction

In the previous two chapters, various interactions of the aromatic moiety in the protein structure and their features in general were described. In the course of these studies, several remarkable phenomena were uncovered deserving a deeper and more detailed investigation. Two of these interactions will be the subject matter of this chapter, divided therefore into two sections.

The first of these phenomena is the interaction between an aromatic residue and a protein backbone, which was encountered in Chapter 3. Analogously to an aromatic group, the NH-CO peptide bond also possesses a conjugated pattern of π orbitals. It has already been mentioned that in rubredoxin the interaction of the aromatic F49 residue with the V5–C6 peptide bond amounted to -8.2 kcal/mol at the CCSD(T)/CBS level, which demonstrates that aromatic residues may interact strongly not only with other sidechains but also with the protein backbone, in the specific case of the XH–π interaction. Due to its extraordinary strength as well as its frequent occurrence (see the discussion in Appendix C), this interaction could play a significant role for the stabilisation of the protein structure. We believe that the origins of this interaction's unexpected strength need to be examined.

The aromatic–peptide bond interaction can occur in two main spatial orientations, analogous to those of the benzene dimer, the stacked and the T-shaped. The π–π stacking may be considered as not only an interaction of aromatic systems but as a broader phenomenon including interactions of planar systems with delocalised orbitals, such as the peptide bond. Similarly, in the XH–π bonding the aromatic ring can serve as an acceptor also for non-aromatic H-bond donors. Unlike in the case of the benzene dimer, where the energetic minima have been described by several extensive and high-level *ab initio* studies, no such investigation has been performed on the aromatic–peptide bond interaction. The few existing studies have employed only the standard computational level. Therefore, it is important to investigate this interaction and describe its basic properties at a sufficiently high level.

In order to find the energy minima for this interaction, geometry optimisations were performed starting from both T-shaped and stacked arrangements, which both occur in proteins. The minimum of the interaction energy was studied using high-level *ab initio* methods and the performance of several other methods evaluated. The major aim of the study was to perform interaction energy decomposition to uncover the underlying forces of this interaction.

The second phenomenon investigated and described in this chapter is the aromatic–proline interaction in the Trp-cage miniprotein, mentioned in Chapter 4. The surprisingly high interaction between the central aromatic tryptophane Trp6 residue with two prolines, Pro17 and Pro18, has already been discussed there. The role of proline in this context – as a source of stabilising interaction – is very interesting, because the classical role of proline residue is an entropical, rather than energetical, stabilisation of the protein native fold. The proline residue has a distinctive position as the only proteogenic amino acid with a secondary α-amino group, and the unique structure of its sidechain provides it with exceptional conformational rigidity. The fact that it contributes less conformational entropy upon unfolding may be one substantiation for the well-known "proline rule", which states that the thermostability of a globular protein can be additively increased by the introduction of proline residues in specific positions. It is also the only amino acid where the cis/trans isomerism of the peptide bond plays a certain role. (For a detailed discussion on the role of the proline, see Appendix D.)

The role of proline in energetical stabilisation has been described by Riley *et al*. for protein–ligand interaction. The authors studied the interaction between the pyrrolidine ring of the proline residue in the binding pocket of the Human Carbonic Anhydrase II and a phenyl ring of its inhibitor. They also suggested that the phenylalanine–proline interaction might be a source of stabilisation for the protein structure, but the strength of such interaction has not been investigated to date.

We decided to study the aromatic–proline interaction on the two specific examples from the Trp-cage structure mentioned above. The question is where this stabilisation originates, *i.e*. what the importance of the cyclic arrangement, the non-aromaticity and the role of the

heteroatom is in the pyrrolidine ring. The MP2 and DFT-D methods were selected for the calculations of interaction energy and the SAPT(DFT) method for its decomposition into interaction energy components. In order to assess the significance of system size and its aromaticity for the interaction, the interaction energy in representative clusters of the Trp-Pro, Phe-Pro, and Pro-Pro pairs extracted from the Atlas of Protein Side-Chain Interactions was also calculated.

5.2 Aromatic–Peptide Bond Interaction

For calculations of the aromatic–peptide bond interaction, the benzene molecule was used as the aromatic partner and three models for the peptide-bond moiety – formamide (FMA), N-methylformamide (NMF) and N-methylacetamide (NMA) (see Figure 5.1).

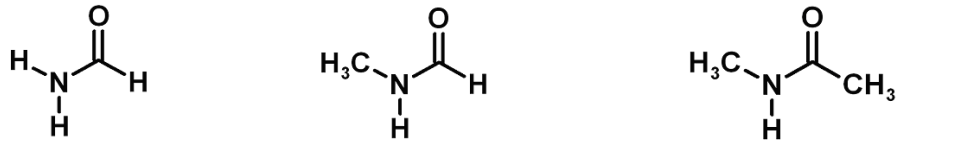


Figure 5.1: The model systems for the peptide bond.

As mentioned above, the interaction may occur in two basic geometrical arrangements: the stacked and T-shaped, which are shown in Figure 5.2 on the example of benzene and FMA. Two geometry optimisations were performed, starting from stacked and T-shaped geometries and the final optimised structures labelled with the suffix “S” and “T”, respectively.

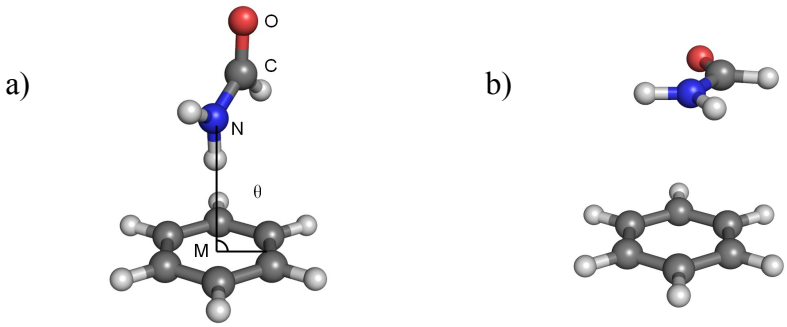


Figure 5.2: The initial geometries of the FMA-benzene complex: a) the T-shaped arrangement; b) the stacked arrangement. The M – the centre of the benzene ring, the θ – the angle between the plane of the benzene ring and the NCO plane of the peptide-bond moiety.

The RI-MP2/cc-pVTZ method was employed for the optimisation of all the subsystems as well as the complexes; the results are summarised in Table 5.1. The optimisation of the T-shape arrangement changes the geometrical parameters of the complex only little and the resulting arrangements are still close to the ideal T-shape geometry. On the other hand, the stacked complexes exhibit a remarkable change during the course of the optimisation, namely as the θ-angle value changes from the initial 0° to values of approximately 50-70°. Since such structures can no longer be called stacked, the term "tilted-T-shaped" was used to distinguish them from both strictly parallel (θ angle 0-20°) and strictly T-shaped structures (θ angle of 70-90°).

Table 5.1: The geometrical parameters of the optimised X-benzene complexes (X=NMA, NMF and FMA) specified by the NM distance between the amide nitrogen of X and the benzene-ring centre (M) and by the interplanar angle θ.

	MP2/cc-pVTZ	
System	**NM [Å]**	**θ angle [°]**
NMA_T	3.221	89
NMA_S	3.188	52
NMF_T	3.210	74
NMF_S	3.144	55
FMA_T	3.274	83
FMA_S	3.259	68

In the optimised complexes, the CCSD(T)/CBS interaction energy was calculated according to the scheme described in Chapter 2, corrected with the deformation energy E^{def}, to obtain the total interaction energy:

$$\Delta E^{total} = \Delta E^{CBS}{}_{CCSD(T)} + \Delta E^{def}. \quad (5.1)$$

All the interaction energies, counterpoise-corrected for the BSSE, are shown in Table 5.2. As expected, stabilisation increases with the size of the basis set from the aug-cc-pVDZ to the CBS by about 1 kcal/mol. It is, however, a known fact that the MP2/CBS values are overestimated relative to the true stabilisation energies. It is therefore necessary to evaluate the CCSD(T) correction term to account for the higher correlation effects. This term brings a stabilisation-energy decrease in all three systems, although it is more profound in the case of NMA and NMF (~1kcal/mol) than in the case of FMA (~0.6 kcal/mol). Overall, the final stabilisation energies, E^{total}, including the deformation energy are very close to those obtained by the MP2/aug-cc-pVDZ method due to the cancellation of errors at this level (basis set incompleteness *vs.* less correlation energy covered).

To summarise, the total interaction energies are relatively high, ranging from -4.6 kcal/mol for the FMA-benzene to -5.2 kcal/mol for the NMA-benzene, which is comparable to classical H-bonding. In accordance with our previous work, where the calculated interaction energy in the phenylalanine–NMF complex was -8.2 kcal/mol, the stabilisation energy rises substantially with the increasing size of the system. It is interesting that the interaction energies of the T-shaped and the tilted-T-shaped arrangements do not differ very much, suggesting that the PES of these systems is relatively shallow and the energy barrier between these arrangements may be low.

Table 5.2: The interaction energies (in kcal/mol) determined for the MP2/cc-pVTZ optimised structures.

System	aDZ	aTZ	CBS	ΔCCSD(T)	CCSD(T)/CBS	E^{def}	E^{Total}
NMA_T	-5.3	-6.0	-6.3	1.0	-5.3	0.3	-5.0
NMA_S	-5.4	-6.1	-6.4	1.1	-5.3	0.1	-5.2
NMF_T	-4.9	-5.6	-5.8	0.9	-5.0	0.0	-5.0
NMF_S	-5.0	-5.7	-6.0	0.9	-5.1	0.1	-5.0
FMA_T	-4.3	-4.9	-5.2	0.6	-4.6	0.0	-4.6
FMA_S	-4.3	-4.9	-5.2	0.6	-4.6	0.0	-4.6

aDZ, aTZ and CBS denote the aug-cc-pVDZ, aug-cc-pVTZ and complete basis set limits, respectively; E^{def} is the deformation energy.

In order to elucidate the source of stabilisation in these interactions, a decision was taken to decompose the total interaction energy using the DFT-SAPT technique. These calculations were performed for the largest of our model systems, the NMA-benzene complex, in the T-shaped and tilted-T-shaped arrangements optimised by the RI-MP2/cc-pVTZ as well as for a set of stacked structures with varying NM distances. Despite no stacked arrangement minimum having been found by our optimisations, such arrangements occur frequently in protein structures, and it will therefore be interesting to include it for comparison.

In the DFT-SAPT calculations, the PBE0AC exchange-correlation functional asymptotically corrected for the description of the dispersion effects was used along with the aug-cc-pVDZ basis set. Although this basis set underestimates the dispersion component by approximately 10-20 %, it should be sufficient for the comparison of relative strengths of this term in the different arrangements. (For computational details, see Appendix D.)

Several facts follow from the results shown in Table 5.3. When the T-shaped arrangement is changed through the tilted-T-shaped into the stacked one with the NM distance kept constant, the electrostatic energy (E^{1}_{pol}) and the effective dispersion term (E^{2*}_{disp}) become more favourable for the interaction. This effect is, however, compensated for by the rising exchange-repulsion term (E^{1}_{ex}). The effective induction energy (E^{2*}_{ind}) is small for all the arrangements.

The differences between the components of the T-shaped and tilted-T-shaped arrangements are tenuous, and the compensation mentioned above yields almost the same total interaction energy in these two arrangements. The differences between the tilted-T-shaped and stacked arrangements are more profound, especially in the E^{1}_{ex} term, which becomes very repulsive in the stacked arrangement due to the enhanced repulsion between the benzene and the NMA methyl groups. This results in much smaller interaction energy in the stacked arrangement (-2.5 kcal/mol *vs.* -7.0 kcal/mol). To reduce this repulsion, the position of methyl hydrogens was optimised and a scan along the NM axis performed. The optimum distance found for the stacked arrangement was 3.5 Å, with an interaction energy value of -5.6 kcal/mol. At this distance, the exchange-repulsion is substantially decreased as is, however, the attractive electrostatic term. On the other hand, the dispersion contribution is comparable to that in the T-shaped and tilted-T-shaped arrangements.

It may be concluded that the interaction energy is comparable for all three arrangements, which is surprising in view of the fact that that H-bonding, commonly seen as the major stabilising feature, is only present in the two T-shaped arrangements. The effective dispersion term constitutes a major contribution to the stabilisation in all three arrangements, but whereas in the T-shaped structures the electrostatic term is almost equally important, in the stacked structure it represents only about one third of the dispersion contribution. The stacked arrangement thus differs from the T-shaped and tilted-T-shaped arrangements in the source of its stabilisation, which is reflected by the significantly higher $E^{2*}_{disp}/E^{1}_{pol}$ ratio.

Table 5.3: The components of the interaction energy (in kcal/mol) determined by the DFT-SAPT for T-shaped (T), tilted-T-shaped (tT) and stacked (S) structures of the NMA-benzene complex.

Arrangement	R(Å)	E^{1}_{pol}	E^{1}_{ex}	$E^{2*\,a}_{ind}$	$E^{2*\,b}_{disp}$	δHF	E_{int}	$E^{2*}_{disp}/E^{1}_{pol}$
T	3.2	-7.3	13.4	-1.8	-10.6	-1.3	-7.1	1.45
tT	3.2	-8.0	15.3	-1.9	-11.6	-1.5	-7.0	1.45
S	3.2	-9.3	24.9	-0.9	-15.3	-1.9	-2.5	1.65
S	3.4c	-4.5	11.0	-0.6	-10.6	-0.8	-5.5	2.36
S	3.5c	-3.4	8.1	-0.5	-9.2	-0.6	-5.6	2.71
S	3.6c	-2.6	6.0	-0.5	-8.0	-0.4	-5.5	3.08

a) $E^{2*}_{ind} = E^{2}_{ind} + E^{2*}_{ex\text{-}ind}$; b) $E^{2*}_{disp} = E^{2}_{disp} + E^{2*}_{ex\text{-}disp}$; c) with the hydrogens of methyl groups optimised.

5.3 Aromatic–Proline Interaction

As mentioned above, two systems were selected for the calculation of aromatic–proline interaction, with the pairs of the residues Trp6–Pro17 and Trp6–Pro18 taken from the structure of the Trp-cage. In order to discover the source of stabilisation for these two contacts, several models of the interacting moieties were employed. In our first model ("Large Model", see Figure 5.3), the molecules were represented in the same way as in the *ab initio* study described in Chapter 4, *i.e.* without their carbonyl group but with the carbonyl of the preceding residue. Subsequently, the tryptophane was reduced to its sidechain starting from the C_β atom and proline represented as a pyrrolidine molecule so that its cyclic structure would be preserved ("Small Model", see Figure 5.3). In order to evaluate the role of the nitrogen heteroatom for the strength of stabilisation, the NH group was replaced with a CH_2 group, thus representing the proline by the cyclopentane model. Last but not least, so as to assess the significance of the cyclic arrangement of proline, an acyclic model was used with the same number of heavy atoms – the C_α model of leucine. The structure of the Trp–Leu pair was taken from the Atlas of Protein Side-Chain Interactions. Only that structure of one of the representative clusters of the Trp–Leu interaction was selected in which the Leu adopts a stacked arrangement relative to the Trp sidechain. (The last two models were used for the Trp6–Pro18 interaction only.)

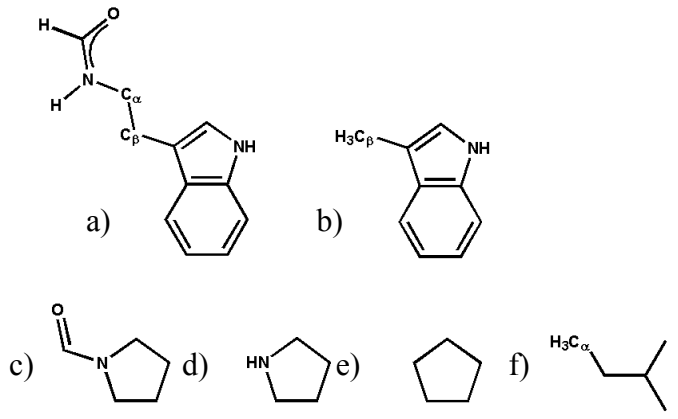


Figure 5.3: The chemical structure of the molecules used as model systems for the aromatic–proline interaction: a) tryptophane – the large model; b) tryptophane – the small model; c) proline – the large model; d) proline – the small model; e) cyclopentane; f) leucine – the C_α model.

In all the studied systems, a geometry optimisation of hydrogens was performed first using the DFT/B3LYP/6-31G** method, followed by the calculation of the interaction energy of the system using the MP2/aug-cc-pVDZ and DFT-D/TPSS/TZVP methods.

The results for the H-bonded Trp6-Pro17 pair are shown in Table 5.4. The difference between the large and small models indicates that the removal of the carbonyl group leads to a decrease in interaction energy by about 6.5 to 7 kcal/mol. We assume that the backbone portion of the Trp residue does not contribute largely to the interaction, being too far from the proline molecule. It is therefore evident that although the interaction energy of this complex is large, it comes almost exclusively from the H-bonding contact. This is also reflected by the dispersion part of the interaction being small and comparable in both of the models employed.

Table 5.4: The interaction energies of the Trp6-Pro17 pair in large and small models using the MP2 and DFT-D methods.

	Trp6-Pro17	
	Large Model	**Small Model**
MP2/aDZ	-7.8	-0.9
DFT-D/tpss/TZVP	-7.6	-1.1
D part	-1.9	-1.1

The situation is more interesting in the case of the Trp6-Pro18 contact, where the two molecules are in a stacked arrangement. The results for this arrangement are summarised in Table 5.5. No H-bond exists between the two interacting systems, and the truncation of the carbonyl group this time leads to only a moderate decrease of the interaction energy. The interaction is in fact dispersion-driven; the dispersion term constitutes a major portion of the interaction energy in both the large and small models. The strength of the interaction is surprising, because it is not a classical stacked interaction of two aromatic systems like the interaction in the benzene dimer, which in the stacked arrangement amounts to -2.73 kcal/mol at the CCSD(T)/CBS level. On the contrary, here the proline molecule is an aliphatic system without any π electrons.

We therefore asked where this large stabilisation energy originates, *i.e.* what the role of the nitrogen heteroatom and the importance of the cyclic arrangement are. To this end, the results for two other models, the cyclopentane and the leucine, were compared. It follows from the data in Table 5.5 that the replacement of the nitrogen heteroatom with carbon reduces the interaction energy by about 1.5 kcal/mol but that resulting interaction energy is still relatively large. The leucine model further demonstrates that the same holds true for the cyclic arrangement of the atoms, because changing the arrangement from cyclic (cyclopentane) to acyclic (leucine) decreases the interaction energy only by about 1 kcal/mol. To sum up, the large stabilisation in the Trp-Pro complex is caused by a number of contributions, none of which clearly prevails.

Table 5.5: The interaction energies in the various models of the Trp-Pro interaction in the stacked arrangement, in the initial geometry and after geometry optimisation calculated using the MP2 and DFT-D methods.

0	**Method**	**Large Model**	**Small Model**	**Cyclopentane**	**Leucine**
Initial	MP2	-8.4	-6.5	-5.1	-4.7
Initial	DFT-D *(D part)*[a]	-6.8 *(-6.3)*	-5.4 *(-5.0)*	-4.0 *(-5.7)*	-3.3 *(-4.6)*
Optimised	DFT-D *(D part)*[a]	-	-11.6 *(-5.0)*	-4.1 *(-5.4)*	-3.6 *(-4.4)*
Optimised[b]	DFT-D *(D part)*[a]	-	-5.5 *(-4.9)*	-	-

a) The "*D part*" denotes the empirically calculated dispersion correction in the DFT-D energy; b) with the position of the N atom restricted.

In order to assess the deviation of the protein geometries of the systems from their energetic minima, a full optimisation in DFT-D/TPSS/TZVP was also performed. The interaction energies calculated using the DFT-D method in the optimised geometries are shown in Table 5.5. Indeed, the geometries changed very little in the cases of the cyclopentane and leucine models (*cf.* the value of interaction energy in the unoptimised complex). On the other hand, the small model complex relaxed into a geometry where an H-bond is formed between the NH group of the Trp and the N atom of the Pro and the resulting enormous interaction energy (-11.6 kcal/mol) is a sign of this additional stabilisation. This H-bond formation is, however, an artefact caused by the fragmentation and does not reflect the possibilities of the proline residue in a protein structure, restricted by the presence of the backbone. If the

geometry optimisation is repeated with the position of the proline's N atom fixed relative to the Trp molecule, the geometry of the complex changes only slightly in the course of optimisation and the interaction energy in this case approaches that of the unoptimised system (see Table 5.5, Column 4). It is therefore concluded that the residue contacts in the protein structures are close to their geometry optima.

One of the questions asked is how large the interaction energy of aromatic–proline interactions is in real proteins. Three interacting systems were the focus of attention - besides the Pro–Trp motif, the Pro–Phe and Pro–Pro pairs were also selected, which represent cyclic systems of aromatic and non-aromatic nature of different sizes. The interaction energies of their representative stacking arrangements obtained from the Atlas of Protein Side-Chain Interactions are shown in Table 5.6.

Table 5.6: Interaction energies of the representative clusters of Trp-Pro, Phe-Pro and Pro-Pro complexes.

	Name[a]	DFT-D *(D-part)*
Trp-Pro	cluster 1	-3.0 *(-5.2)*
	cluster 2	-3.7 *(-3.8)*
	cluster 3	-3.3 *(-5.1)*
	cluster 4	-3.4 *(-3.4)*
	cluster 5	-3.2 *(-6.0)*
Phe-Pro	cluster 1	-2.3 *(-2.9)*
	cluster 2	-2.7 *(-2.8)*
	cluster 3	-2.4 *(-3.9)*
Pro-Pro	cluster 1	-1.6 *(-2.7)*
	cluster 2	-2.3 *(-2.5)*
	cluster 3	-0.7 *(-1.7)*
	cluster 4	-2.2 *(-2.9)*

a) The sequence number in the Atlas of Protein Side-Chain Interactions.

It is quite clear that the interaction energy rises with the size of the interacting systems, *i.e.* Pro–Pro < Phe–Pro < Trp–Pro. The highest interaction energies reached values of -2.3 kcal/mol for Pro–Pro, -2.7 kcal/mol for Pro–Phe and -3.7 kcal/mol for Pro–Trp. The average values varied between -1.7 kcal/mol for Pro–Pro to -3.3 kcal/mol for Pro–Trp. As arises from Table 5.6, it is the dispersion which is chiefly responsible for such strong stabilisation. Evidently, the cluster-representative geometries can be quite far from the actual geometry of

the residues in a real protein and the values thus do not represent the most ideal arrangement of the interaction. In the view of this fact, it is noteworthy that the Pro–Trp interaction reaches values of a classical hydrogen bond in all the cases.

5.4 Conclusions

This chapter has focused on two specific stabilising interactions in proteins, namely the aromatic peptide bond and the aromatic–proline interaction. These findings clearly demonstrate that not only the interactions of aromatic residues in the sidechain–sidechain manner but also their interactions with the protein backbone can be a significant source of protein stabilisation.

The aromatic ring–peptide bond interaction was studied by several advanced computational chemistry methods. The MP2 geometry optimisations revealed two geometry minima, the T-shaped and the tilted-T-shaped arrangements, which are very close in energy. The interaction energies in both arrangements are large and reach values lower than -5 kcal/mol at the CCSD(T)/CBS level (the NMA-benzene model), which is comparable to a strong H-bond. Given the high occurrence of these contacts in the protein structure, the contribution to the protein stabilisation can be significant.

The employment of the SAPT technique enabled the decomposition of the interaction energy into its components and the elucidation of the different nature of the stabilisation in the stacked and the T-shaped arrangements. The two arrangements, despite having similar interaction energy, differ greatly in the proportion of electrostatic and dispersion components.

In the case of the aromatic–proline residue interactions, which represent a crossover between the classical sidechain–sidechain and the sidechain–backbone interactions due to the specific character of the proline residue, a series of models were employed in order to trace the origin of the stabilisation. The study revealed that both the nitrogen heteroatom of the pyrrolidine ring and the favourable cyclic arrangement contribute to the surprisingly high interaction

energy in this interaction. The interaction strength also depends on the size of the aromatic moiety. It is the dispersion component that is chiefly responsible for the stabilisation.

4 Interaction Energy Matrices in Proteins

4.1 Introduction

The preceding chapter discussed the strength of interactions maintained mostly by aromatic residues composing the protein hydrophobic core. It was clearly demonstrated that these interactions are strong and that the role of aromatic residues in the stabilisation of protein structure can be significant. So as to be able to evaluate their role more precisely, it is necessary to know their relative strength when compared to the interactions of other amino acids in the protein structure. The aim of the chapter is to study the role of aromatic residues in a wider structural context.

The strategy is to calculate mutual pairwise interactions of all residues in a protein structure, splitting the polypeptide chain into fragments corresponding to the individual residues and presenting the residue–residue interactions in a square matrix. This matrix may resemble the frequently used contact map representation of protein structure, which shows the distances between all residue pairs (usually their C_α atoms). Unlike the contact map, however, the Interaction Energy Matrix (IEM) introduced here captures the topological relationships between the residues in a protein better and makes it possible to recognise the patterns which are of key importance for the stabilisation of the structure. In fact, a similar concept has previously been used by Vendruscolo *et al*. when examining the structure of a transition state ensemble (TSE) in the protein folding process. Using constrained MD simulations, the authors showed that a small number of residues is sufficient to determine the native-like topology in the TSE. These "key residues" are characterised by a high $\langle\phi\rangle_i^{exp}$ value (the ratio of the change in the stability of the TSE to that of the native state caused by mutation of residue *i*). The key residues typically correspond to the residues of a hydrophobic core and form a large number of stabilising contacts with other residues.

Several methods have been proposed for the determination of the key residues in the native structure of protein, for example mutational studies and sequence conservation, the Gaussian network model analysis or the distance-based contact theory. None of these, however, has proved to be both reliable and easy to conduct (for a detailed discussion, see Appendix B). It has been proposed by Chen and Xiao that the IEM strategy could be used for the

determination of the key residues. The authors used the CHARMM27 force field with the GB/SA solvent model on a G_β domain of a multi-domain protein transducin. In this case, however, the residues that were identified as the most strongly interacting and therefore the key ones were mainly the charged residues that link together the blades of the propeller-like protein. We believe that the situation could be very different in a small globular protein, because this inter-subdomain stabilisation would not be present. Therefore, in addition to calculating all inter-residue interactions in a small globular protein, the aim of this chapter is to investigate how the relative strength of different interactions is related to the possible identification of some residues as the key ones in structure stabilisation.

4.2 Methods and Models

On the basis of our experience gained in the study of the rubredoxin core (Chapter 3), the RI-MP2/aug-cc-pVDZ was selected as the reference method for this study. Such a method is, however, very computationally demanding and feasible only for a small model system. If the IEM concept is to be applied routinely in large proteins, faster methods are required. Two methods were thus tested: the DFT-D/TPSS/TZVP method, which can also describe the dispersion part of the interaction, and the parm94 force field of the Amber package, which is often used in protein simulations.

As different noncovalent interactions between the residues in proteins are affected by the environment to a different degree based on their nature,[*] it was necessary to take the role of the environment into account. In order to describe the environment in the DFT-D calculations, the implicit solvent model (COSMO) was employed with the value of the dielectric constant $\varepsilon = 80.0$ and other constants kept at their default values as implemented in the Turbomole program package. The Generalised Born (GB) solvent model was used with Amber calculations with the dielectric constant $\varepsilon = 78.5$, the standard value for water. Taking into consideration that the dielectric constant of the protein interior is much lower than that of bulk water (usually values between 4.0 and 10.0 are used), the present values have been

[*] For example, the charge–charge interactions are more shielded by the dielectric continuum than the dispersion ones.

selected so that the resulting interaction energy values will represent the lower limit of the actual interaction strength for both solvent-exposed and buried residues.

Since the number of interaction rises proportionally to the square of the number of residues in a protein, it was necessary to choose a sufficiently small model system in order to be able to calculate the whole inter-residue interaction energy matrix using the high-level reference MP2 method. The Trp-cage miniprotein (PDB code 1L2Y), the smallest (20-residue) artificial protein-like structure known to date, was selected (see Figure 4.1 for the Trp-cage structure). The aromatic trypthophane residue forms the centre of the hydrophobic core of this protein and participates in a number of contacts with structurally adjacent residues.

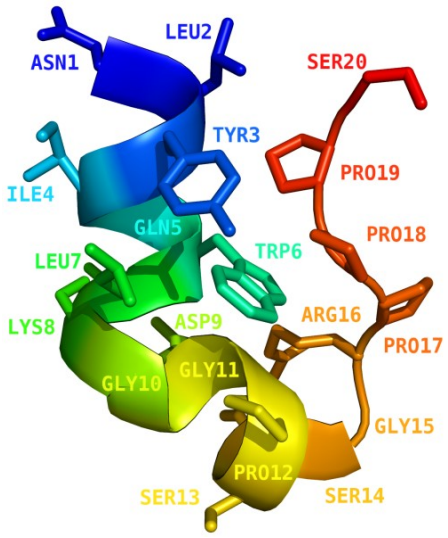


Figure 4.1: The structure of the Trp-cage miniprotein, PDB code 1L2Y.

In order to be able to calculate the pairwise interaction energy between the residues, the polypeptide chain of the protein has to be split into separate amino acid (AA) residues. Different fragmentation procedures were used for the *ab initio* (MP2 and DFT) and the Amber calculations. So as to preserve the important peptide bond of the backbone, the C_α-C(O) bond of the backbone was split and the resulting fragments were capped with H atoms in the *ab initio* calculations (see Figure 4.2). For the Amber calculations, the peptide bond between the residues was cut and the fragments capped with $H_3C–C=O–$ and $–NH–CH_3$ groups. (Note that this scheme also renders the peptide bond moiety but makes the systems

much larger, see Figure 4.2.) These schemes enabled the calculation of the interaction energy in all the pairs of non-neighbouring residues in the case of the *ab initio* methods, or in all the pairs of *n*-th and *m*-th ($m > n + 2$) residues in the case of the empirical method. In both types of calculations, the positions of the hydrogens were optimised prior to the calculation of interaction energy (see Appendix B for details). For a graphical illustration, the interaction energy values for all pairs are gathered in a matrix (see Figure 4.3 a) for an example).

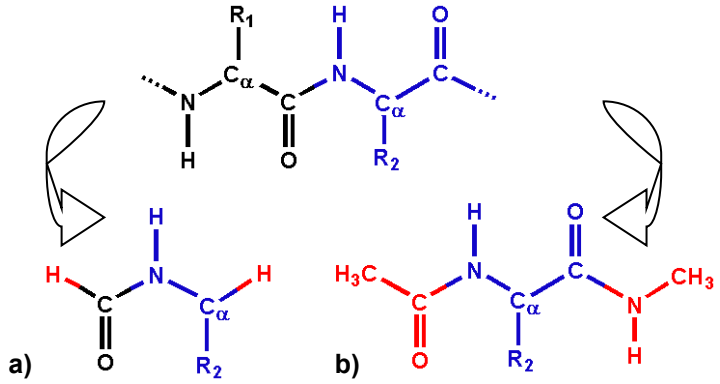


Figure 4.2: The fragmentation schemes used: a) with MP2 and DFT-D methods; b) in Amber calculations. The selected residue is shown in blue, the preceding residue in black, the groups used for fragment capping in red.

4.3 The Interaction Energy Matrix for the Trp-cage Miniprotein

The gas-phase calculations were first performed at the reference MP2 level. As expected, the interactions both between charged residues and between charged and neutral residues were strongly overestimated because of the unshielded Coulomb part of interaction in the gas phase (see Appendix B for details). With respect to the fact that most of the residues necessary for Trp-Cage stability are neutral,* the decision was taken to omit the charged residues from the gas-phase analysis (the residues Lys8, Asp9 and Arg16, as well as the charged termini, Asn1 and Ser20). The resulting IEM is shown in Figure 4.3.

* The consensus sequence between the various Trp-cage proteins synthesised is X**F**XX**W**XXXX**GP**XXXX**PPP**X. A ConSeq analysis of the exendin-4 fragment, which was used to design the Trp-cage, has revealed the residues F3, V/I4, W6, L7, G10 and P12 to be the most conserved.

An important finding is that the interaction energy of most contacts is attractive; wherever repulsion occurs, it is very small or negligible. Moreover, several contacts, especially the interactions of the aromatic residues Tyr3 and Trp6, exhibit very large stabilisation. The Tyr3 interacts most strongly with Trp6, and the energy content of this interaction is about -8.5 kcal/mol. This interaction is partly due to the backbone–backbone H-bond between these two residues, which stabilises the alpha helix. The Trp6 residue, having a large aromatic sidechain, possesses several other extremely strong interactions. Besides the interaction with Tyr3, Trp6 interacts strongly with the prolines Pro17 and Pro18. The interaction between Trp6 and Pro17 (of -7.5 kcal/mol) is stabilised by an H-bond between the nitrogen of the tryptophane indole ring and the carbonyl oxygen of Pro17. On the other hand, there is no H-bond between the interacting residues Trp6 and Pro18. The interaction (of -8.7 kcal/mol, even stronger than between Trp6 and Pro17) originates solely in stacking. This is surprising, because one of the interacting partners has not an aromatic but an aliphatic ring and no CH-π interaction is involved. Both of these interactions will be discussed in further detail in the following chapter.

In addition to comparing the strength of individual interactions, it is also useful to compare the total interaction energy brought by a single residue (shown in Figure 4.3 b). The total contribution of all residues is stabilising, and indeed most of the stabilisation is brought by the interactions of the central aromatic Trp6 residue. The contributions of the residues Leu7, Tyr3 and Gly11 (all of them larger than -20 kcal/mol) are also important. This set shows a successful overlap with the consensus sequence of Trp-cage variants (see above). The conclusion thus is that this method could be used for the identification of key residues in proteins.

a)

	2	3	4	5	6	7	10	11	12	13	14	15	17	18	19	
2			-5.1	-6.7	-1.3	-0.6	-0.1	-0.1	0.0	-0.1	0.0	0.0	0.1	-0.1	-0.4	2
3				-4.7	-8.5	-3.0	0.0	-0.4	0.3	0.1	-0.2	0.1	0.0	-0.8	-3.7	3
4	-5.1				-4.8	-7.5	-0.4	-0.6	0.0	-0.1	-0.2	0.0	0.3	-0.1	-0.4	4
5	-6.7	-4.7				-3.6	-0.8	-0.4	-0.1	-0.2	-0.1	0.0	0.2	-0.2	0.2	5
6	-1.3	-8.5	-4.8				-0.8	-3.8	-5.2	-0.9	-0.3	0.0	-7.5	-8.7	-3.3	6
7	-0.6	-3.0	-7.5	-3.6			-2.1	-6.8	0.8	-0.1	-1.2	0.2	0.6	0.2	0.1	7
10	-0.1	0.0	-0.4	-0.8	-0.8	-2.1			0.2	-0.2	-3.4	0.2	-0.1	0.1	0.1	10
11	-0.1	-0.4	-0.6	-0.4	-3.8	-6.8				-6.3	-1.0	-0.6	-0.4	-0.2	0.2	11
12	0.0	0.3	0.0	-0.1	-5.2	0.8	0.2				-6.2	-0.4	1.7	0.5	-0.2	12
13	-0.1	0.1	-0.1	-0.2	-0.9	-0.1	-0.2	-6.3				-3.7	0.3	0.0	0.1	13
14	0.0	-0.2	-0.2	-0.1	-0.3	-1.2	-3.4	-1.0	-6.2				-1.5	-0.2	0.1	14
15	0.0	0.1	0.0	0.0	0.0	0.2	0.2	-0.6	-0.4	-3.7			-0.5	0.3	-0.1	15
17	0.1	0.0	0.3	0.2	-7.5	0.6	-0.1	-0.4	1.7	0.3	-1.5	-0.5			-2.2	17
18	-0.1	-0.8	-0.1	-0.2	-8.7	0.2	0.1	-0.2	0.5	0.0	-0.2	0.3				18
19	-0.4	-3.7	-0.4	0.2	-3.3	0.1	0.1	0.2	-0.2	0.1	0.1	-0.1	-2.2			19
	2	3	4	5	6	7	10	11	12	13	14	15	17	18	19	

b)

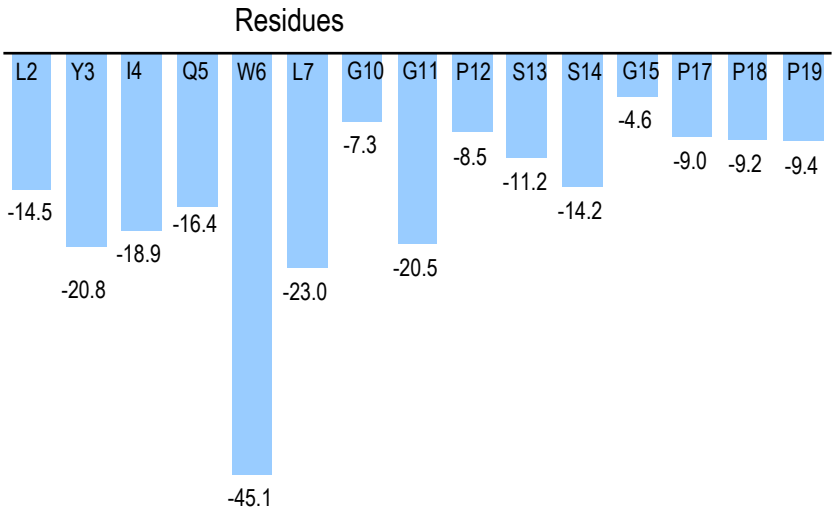


Figure 4.3: The MP2 gas-phase IEM for the uncharged residues: a) the individual pairwise interactions; b) the total interaction energy for each residue as a sum of its pairwise interactions. All values are in kcal/mol. The dark blue: (-∞) – (-5.0) kcal/mol, the light blue: (-5.0) – (-0.5) kcal/mol, the grey: (-0.5) – 0.5 kcal/mol and the yellow: 0.5 – ∞ kcal/mol.

The next step was the evaluation of the performance of the two methods tested, the DFT-D and Amber empirical potential. A comparison of the total interaction energies brought by the individual (uncharged) residues as calculated by the reference MP2 as well as the two tested methods is shown in Figure 4.4. The DFT-D method provides excellent agreement with the MP2. On the other hand, the results of the Amber method were overestimated on average by a factor of 2.0 (the correlation coefficient R = 0.842) as compared to the MP2 values. This is mostly caused by the different fragmentation method used in the Amber calculations, which produces larger fragments and thus leads to larger stabilisation energies due to the increased dispersion interaction. After rescaling the Amber values by a factor of 2.0, they reproduce the values of the reference method relatively well. An important finding is that all three methods identify the same set of residues (*i.e.* Trp6, Tyr3, Leu7 and Gly11) as the key ones for the structure stabilisation.

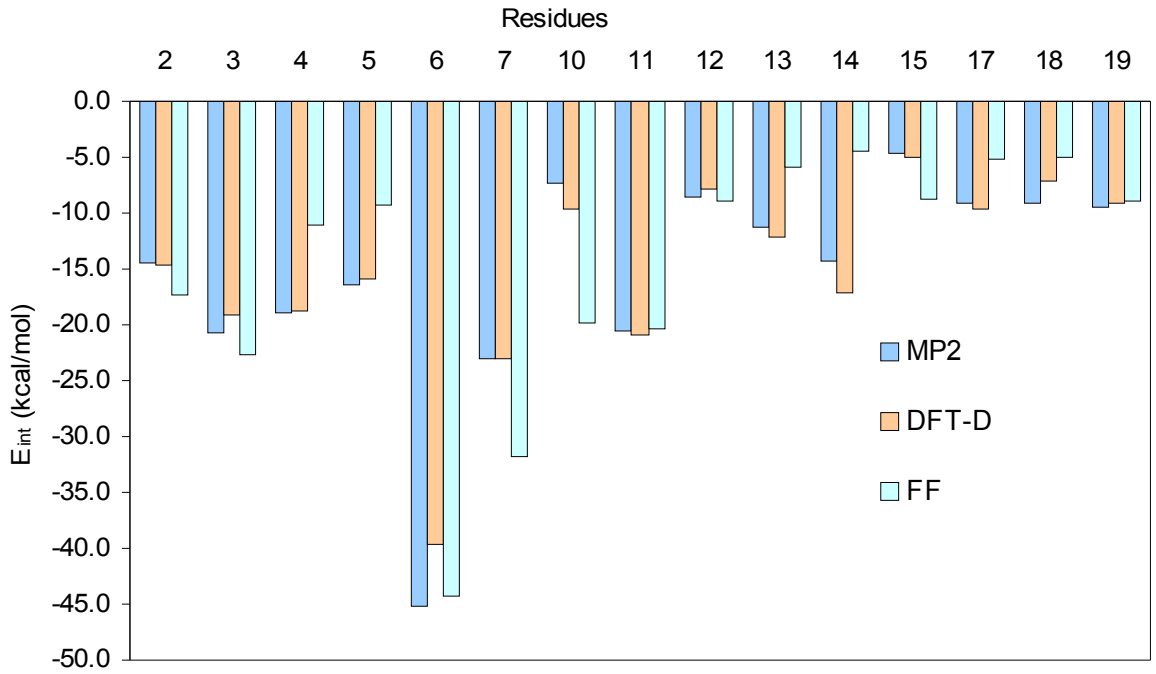


Figure 4.4: A comparison of the total interaction energies for the individual Trp-cage residues in the gas phase as calculated by the MP2, DFT and Amber methods (in kcal/mol). The Amber values have been scaled down by a factor of 2.0.

After the evaluation of the inter-residue interaction energies in the gas phase, the focus shifted to the evaluation of the influence of the environment. Because of the excellent agreement between the MP2 and DFT-D methods in the gas phase calculations and the high computational demands of MP2, the DFT-D was used as the reference method in order to test

the Amber potential. As follows from the results shown in the graph in Figure 4.5, the contributions of most residues are stabilising when measured by means of the reference DFT-D method. The contribution of the charged residues (especially of the residue Lys8) is still high but in general is substantially diminished from the gas-phase values (see Appendix B for details). The interactions of the nonpolar residues are also decreased but by only approximately one half of their gas-phase value. The aromatic Tyr6 remains the strongest stabilising residue. The prominent role of the residues Tyr3, Leu7 and Gly11 is evident, with the set of residues having a total interaction energy of over -6 kcal/mol being the residues Tyr3, Gln5, Trp6, Leu7, Lys8, Gly11 and Pro19.

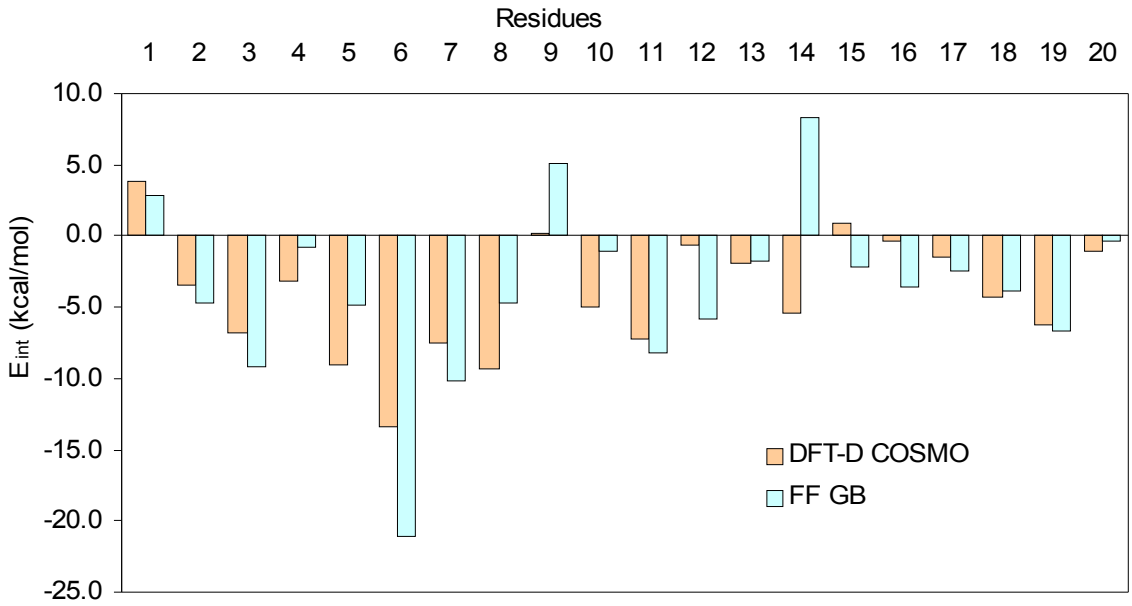


Figure 4.5: A comparison of the total interaction energies for the individual residues as calculated with the solvent model by the DFT-D/COSMO and Amber/GB (in kcal/mol). The Amber values have been scaled down by a factor of 1.4.

The values calculated by the Amber/GB method were again systematically overestimated (when compared to this case's reference DFT-D), but the scaling factor this time was 1.4 (with the correlation coefficient R = 0.67). After rescaling, the agreement between the two methods is relatively good. It identifies the same set of key residues (having the total interaction stronger than -6 kcal/mol), with the only exception being the residue Pro12 identified instead of Lys8, which again might be a result of the different fragmentation procedure. Despite this slight inaccuracy, we believe that the method has enough predictive

power to be used for the identification of key residues in large proteins, where it is necessary to use a computationally inexpensive method (for examples of the applications, please refer to Appendix B).

4.4 Conclusions

The focus of this chapter has been the role of aromatic residues in the protein structure in terms of the pairwise interactions between all protein residues while utilising the pairwise inter-residue IEM. At the same time, the IEM was also examined as a new and useful tool for analysing the protein structure, characterising not only the protein fold (analogically to the more common contact map) but also the energetic relations between the residues inside the protein. In order to find a reliable and at the same time sufficiently fast method which could be used also for large proteins, the MP2 method was employed as a reference and compared with the performance of two other methods, the DFT-D and Amber empirical potential.

In the gas-phase calculations of uncharged residues, the agreement between the MP2 and DFT-D was exceptional, and the performance of the Amber (after scaling) was also sufficiently good. When the environment effects were included, a qualitative agreement was found between the DFT-D/COSMO and Amber/GB methods. The conclusion is that both methods are robust enough to be used in connection with the IEM for the localisation of the key residues responsible for protein structure stabilisation.

The aromatic residues have been shown to play a prominent role for the stabilisation of the structure of our model protein, Trp-cage. Both the aromatic residues present in the structure, Tyr3 and Trp6, were identified as key by all the calculations. Not only do these residues interact by strong aromatic–aromatic interactions, but also their aromatic–aliphatic interactions (*e.g.* Trp6-Pro18) can be exceptionally large. A specific example of this aromatic-aliphatic interaction, the interaction between the aromatic residue and proline, will be the focus of Chapter 5.

3 Aromatic Interactions in the Protein Core

3.1 Introduction

The folding process in which a protein acquires its unique three-dimensional structure is characterised by two crucial features – stability and specificity. Stability means that under the native conditions the folded conformation has lower free energy than the unfolded one, whereas specificity ensures that there is only one such native structure. As Anfinsen's principle states, the fold of a protein is encoded in its amino acid (AA) sequence. The native fold is determined by the sequence of a protein through a large multitude of short-range noncovalent interactions among AA residues close in sequence (secondary structure) as well as long-range noncovalent interactions among sequentially distant residues (tertiary structure). Bearing in mind that the folding free energy of a typical protein is only about -10 kcal/mol, it is obvious that even very small differences in interaction energy may play an important role in this well-balanced system.

The noncovalent interactions stabilising the native fold include hydrogen bonding, electrostatic, induction and dispersion forces. In order to be able to engineer protein structure and function in a targeted way, a deep understanding is necessary of how the fine balance between these noncovalent interactions is achieved. While the importance of ionic and H-bonded interactions is well recognised, the opposite holds true about the role of dispersion energy controlled interactions. The dispersion interactions are numerous but had been long believed to be considerably weaker than the polar interactions (such as H-bonding) and thus less important.

The very first step in the process of protein folding is a hydrophobic collapse of the aliphatic/hydrophobic AAs into the interior of the protein. The collapse, which is the driving mechanism of protein folding in terms of Gibbs free energy, is believed to be an example of the classical hydrophobic effect dominated by the entropy contribution of the solvent molecules, with the enthalpy contribution of the hydrophobic sidechains complexation being small.

The fact that the core of a typical protein has only few H-bonds leads to the assumption that the enthalpy contribution of the entire core formation to protein folding would be negligible.

Nevertheless, it has become apparent from recent theoretical as well as experimental investigations that not only H-bonding but also other types of noncovalent interactions can be characterised by relatively large interaction energy, for example the van der Waals interactions in the benzene dimer or in the stacked DNA base pairs.

As for the aromatic-aromatic interactions in proteins, the distinct packing arrangements of two phenylalanine residues, analogical to the crystal packing of benzene molecules, were described by Burley and Petsko already in the 1980s. Using the benzene dimer model and a simple empirical potential, the authors also estimated the interaction energy to be between -1 and -2 kcal/mol for a typical phenylalanine pair, suggesting its role in protein-structure stabilisation. This method, however, could not take the precise role of dispersion interactions into account, and therefore the true interaction energy might be much higher.

Dispersion interactions are characterised by a shallow PES. In other words, the interactions are very nondirectional, unlike the H-bond, which allows the aromatic residues to participate in a large number of interactions. This property is also demonstrated in the simulations of the transition state ensemble (TSE), which have revealed that the small number of residues which are necessary to define the native-like topology of the folding TS via their tertiary contacts typically correspond to the hydrophobic core.

Taking into account the above-mentioned facts along with the presence of a non-negligible fraction of aromatic residues in the sequences and structures of proteins, it is evident that a thorough examination of the interaction between the residues constituting the protein core is needed. The aim of this chapter of the thesis is therefore to study the strength of the noncovalent interactions of those aromatic residues forming a hydrophobic cluster using appropriate methods, capable of describing the important dispersion part of the interaction correctly. This will make it possible to answer the question how strong the interaction of residues in the hydrophobic core is and consequently how significant their contribution to the protein native fold stabilisation is.

3.2 The Model Case of Rubredoxin

The model system for this study, the small protein rubredoxin, was selected based on the following criteria: it needed to be a globular single-domain protein of up to 100 residues with a well-defined hydrophobic core with a high resolution X-ray structure available. (For details

on the selection as well as the whole study, see Appendix A.) These criteria made it possible to calculate the interaction energies of the aromatic and hydrophobic residues in the protein core in their biologically relevant conformations. Fulfilling these criteria, rubredoxin (PDB code 1RB9), a 52-residue protein with a resolution of 0.92 Å, was selected from the PDB database. The hydrophobic core of rubredoxin is centred around two phenylalanine residues, F30 and F49, as shown in Figure 3.1.

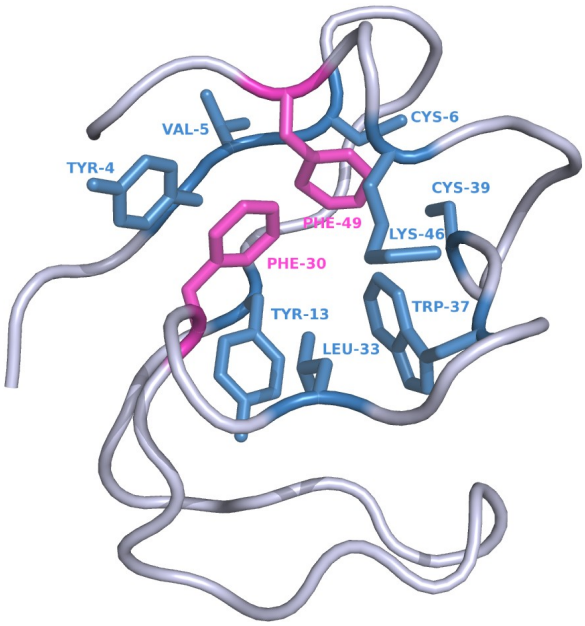


Figure 3.1: The rubredoxin protein with the hydrophobic cluster centred around the residues F30 and F49 shown.

The localisation of the core was carried out using STING Millennium, a web-based suite of programs, while employing a distance criterion of 4.2 Å for hydrophobic interaction and selecting the residues with the most interacting partners as the central residues of the hydrophobic cluster. The hydrophobic core of rubredoxin, centred around the residues F30 and F49 as mentioned above, can be divided into two subclusters of residues, each interacting with one of the central phenylalanines. In the F30 subcluster, the residues Y4, Y13, L33, K46 and F49 interact with the central phenylalanine F30, after which it has been named. In the F49 subcluster, the residues Y4, V5, C6, F30, W37, C39 and K46 interact with the central phenylalanine F49. The structures of the subclusters are shown in Figure 3.2. In both subclusters, most residues interact in hydrophobic interactions; there is only one H-bond

located in the F30 subcluster (a classical CO–HN H-bond in the F30–L33 pair) and two H-bonds in the F49 subcluster (a classical CO–HN H-bond in the F49–K46 pair as well as an unusual CH–π interaction between the methyl group of the capped O terminus of V5 and the π system of the phenylalanine in the F49–V5 pair).

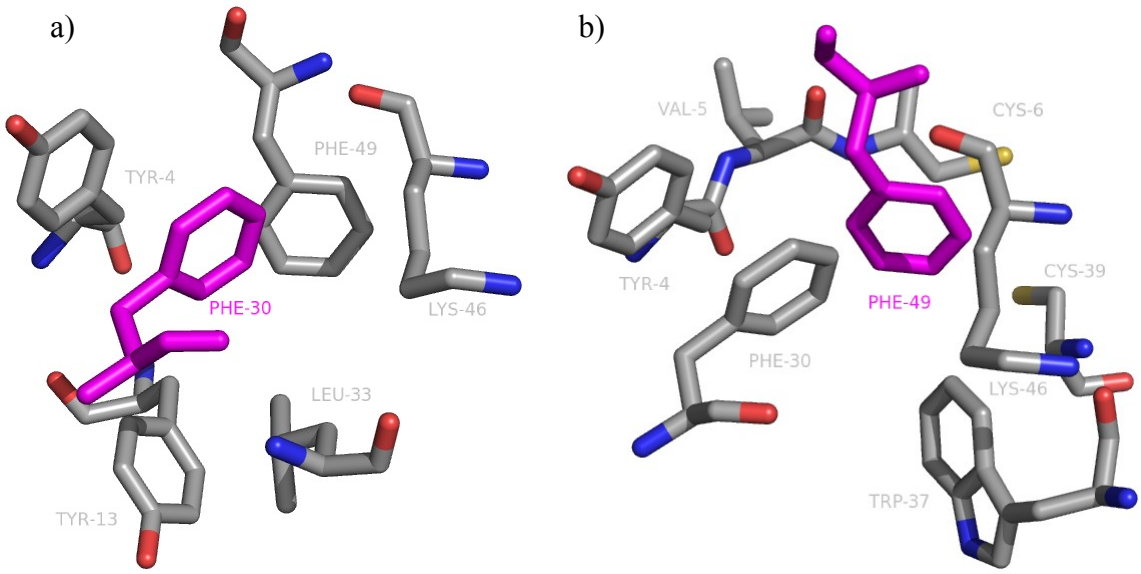


Figure 3.2: a) the F30 subcluster, b) the F49 subcluster

Both subclusters were fragmented into chemically distinct pairs of interacting AAs (modelled as methylated aminoacyl residues). The heavy-atom coordinates in all AA pairs remained fixed at the positions from the X-ray structure. The positions of hydrogens were determined by means of the DFT/B3LYP/ 6-31G** optimisation of the particular pair. The stabilisation energies for all pairs of AAs were determined at the RI-MP2/CBS level; extrapolation to the CBS limit was done according to the scheme of Halkier *et al*. on the basis of aug-cc-pV*x*Z (*x* = D, T) basis sets. For several selected pairs, the CCSD(T)/CBS level interaction energy was calculated as a benchmark according to the protocol described in Chapter 2. For the sake of comparison, the interaction energies in the commonly used B3LYP functional were calculated with the 6-31G** basis set. All interaction energies were CP-corrected for the BSSE.

The interaction energies of all AA pairs are summarised in Table 3.1. All interaction energies are non-negligible at the RI-MP2/CBS values and vary between -2.1 and -7.0 kcal/mol, the largest interaction energy having been ascertained for the F30–Y4 pair, a purely dispersion-

driven interaction. The largest interaction energy was found for the F49–V5 pair, namely of -7.0 kcal/mol. The total interaction energy was determined as the sum of the pair-wise interaction energies of all AAs with the respective central phenylalanine.

Table 3.1: The interaction energies (in kcal/mol) in the F30-F49 cluster evaluated by *ab initio* (RI-MP2, CCSD(T), DFT), semiempirical (DFTB-D) and empirical (CHARMM and Amber) methods.

		aDZ[a]	aTZ[b]	CBS[c]	Δ[d]	CCSD(T)[e]	B3LYP	DFTB-D	CHARMM	AMBER
F30 subcluster	**F49**	-3.1	-3.3	-3.3	-	-	0.8	-3.0	-2.1	-19.3
	K46	-3.1	-3.3	-3.4	0.3	-3.1	0.0	-2.9	-1.3	-13.0
	L33	-4.9	-5.3	-5.5	0.5	-5.0	1.8	-4.6	-2.7	-6.6
	Y13	-4.2	-4.4	-4.5	0.6	-3.9	0.2	-3.7	-2.4	-3.1
	Y4	-6.5	-6.8	-7.0	-	-	3.2	-6.3	-4.7	-5.6
F49 subcluster[f]	**C39**	-1.7	-2.0	-2.1	-	-	1.2	-2.2	-1.1	-7.4
	C6	-4.4	-4.8	-5.0	-	-	2.6	-5.2	-2.5	-17.1
	K46	-4.0	-4.6	-4.8	-	-	0.2	-4.5	-2.6	-11.2
	V5	-5.6	-6.4	-6.7	-	-	1.6	-7.1	-1.4	-8.1
	W37	-2.3	-2.4	-2.5	-	-	0.7	-2.0	-1.5	-4.0
	Y4	-2.7	-3.0	-3.1	-	-	2.8	-4.3	-2.6	1.4
	Total	-42.5	-46.4	-47.9	-	-	15.1	-45.8	-24.8	-93.8

a) RI-MP2/aug-cc-pVDZ, b) RI-MP2/aug-cc-pVTZ, c) RI-MP2/CBS limit, d) ΔCCSD(T)/6-31G*(0.25) e) CCSD(T)/CBS, f) F49 subcluster without the F30 residue.

It is surprising that all the examined interactions exhibit negative and relatively strong interaction energy (for six pairs even stronger than -4.5 kcal/mol, see Table 3.1) at the RI-MP2/CBS level. In the traditional view of protein-fold stabilisation, the main contribution comes from the H-bonds and the entropy contribution of the solvent upon the aggregation of the hydrophobic sidechains. Therefore, the interactions of the aromatic and aliphatic residues are expected to be repulsive or only slightly attractive. Nevertheless, the total interaction energy in the whole selected cluster amounts to nearly -50 kcal/mol (with the interaction energy of the F30 residue with the surrounding AAs amounting to nearly -24 kcal/mol and the interaction energy of F49 residue to nearly -28 kcal/mol).

In order to be able to estimate the contribution from the higher-order correlation terms, the Δ CCSD(T) correction was determined for three selected residue pairs: F30–K46, F30–L33, and F30–Y13. This made it possible to approximate the CCSD(T)/CBS interaction energies for these three pairs. In all three cases, the ΔCCSD(T) correction term, whose calculation is computationally very demanding, is relatively small. Therefore, it may be assumed that the

RI-MP2/CBS level is sufficient for the correct description of the studied systems. The interaction energies determined at the frequently used DFT/B3LYP/6-31G** level are either repulsive or negligible for all the residue pairs investigated (see Table 3.1 and Figure 3.3). This picture, despite being consistent with the traditional view of the stabilisation forces inside a protein hydrophobic core, is incorrect. This has been indicated by a comparison with more advanced methods, *e.g.* MP2 (see Figure 3.3). The reason why the DFT is not able to describe the strong attraction between the hydrophobic side chains of these AAs correctly is mainly the poor description of dispersion energy by the DFT functionals. It is therefore necessary to emphasise the inappropriateness of this method for calculations of all the systems where van der Waals attraction is dominant.

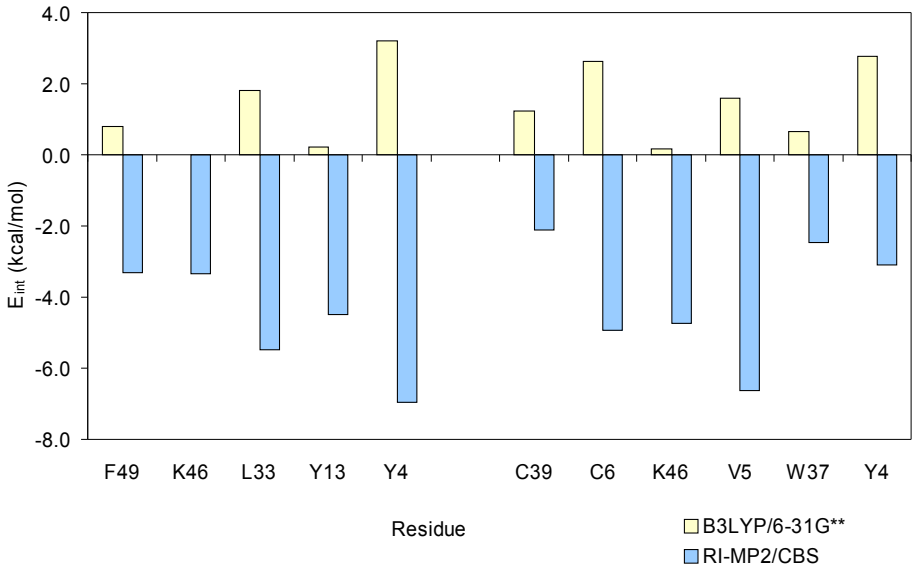


Figure 3.3: The interaction energies in the F30 and F49 subclusters determined at the RI-MP2/CBS and DFT/B3LYP/6-31G** levels.

A comprehensive examination of the particular interaction pairs reveals that there are two especially strongly interacting pairs, namely the F30–Y4 pair and the F49–V5 pair, both with interaction energies stronger than -6.5 kcal/mol (for a more detailed description of these interesting residue contacts, see Appendix A). These examples indicate that a rather strong stabilisation can originate not only from the classical H-bonding interaction but also from the less specific van der Waals interactions. In fact, the interaction energy is significant in all 11 residue pairs and amounts to -4.4 kcal/mol on average, a value comparable to the strength of a classical H-bond, far from the expected negligible or even repulsive one. Therefore, the

stabilisation of the protein core is significantly facilitated by not only aromatic-aromatic interactions but also the aromatic-aliphatic interactions of the central phenylalanine and all the neighbouring residues. The fact that the aromatic residues are able to participate in a large number of stabilising contacts simultaneously can have significant implications for the mechanism of protein folding, where these residues can play a role of connecting hubs or key residues, which will be discussed in more detail in Chapter 4.

In this case, the total interaction energy in the protein core has been estimated to nearly -50 kcal/mol. In order to be able to compare this value with the average folding free energy of a protein, of about -10 kcal/mol, several factors would have to be considered. First, the entropy contribution, desolvation and cavitation energies would all have to be taken into account and the influence of the environment on the strength of the interactions would have to be accounted for. The dielectric screening of the protein interior influences different interaction types differently, with the largest effect occurring in the case of charged interactions and the hydrophobic (dispersion) contribution being the least affected. However important these effects might be, it is clear that such strong stabilisation has to have an enormous impact on the protein stability as well as the folding process itself. It is clear that the view of protein folding and tertiary structure stabilisation as a solely entropy-driven process has to be modified. The present study proves that the energy contribution of the core formation is significant and should not be overlooked. These conclusions have also been corroborated by a combined computational and calorimetric study of mutants of a hyperthermophilic variant of rubredoxin.

3.3 Conclusions

The *ab initio* calculations (RI-MP2/CBS) presented in this chapter have proved that the protein hydrophobic core is stabilised by significant dispersion-type interactions of an aromatic–aromatic as well as aromatic–aliphatic nature, which are on average comparable to the strength of a classical H-bond.

The ΔCCSD(T) correction term evaluated for selected residue pairs is small enough for the RI-MP2/CBS level to be considered as sufficient for the studied systems. It has also been confirmed that the B3LYP functional is not suitable for such systems, due to its failure when describing the London dispersion forces.

The total interaction energy in the core estimated at the RI-MP2/CBS level is large and amounts to nearly -50 kcal/mol. Hence the current view of the hydrophobic core formation as a consequence of the entropy-ruled process should be revised.

2 Computational Methods

An important step in the theoretical study of any molecular system is an astute selection of appropriate computational methods. The two most important issues to come into account here and necessary to be carefully considered are the precision of the method versus the computational cost. In this chapter, the properties of various computational procedures applicable in noncovalent interactions of aromatic systems in their ground electronic state will be discussed.

The main features of an interaction between two moieties that can be described computationally are their mutual geometry arrangement and the interaction energy, which quantifies the strength of the interaction for the particular spatial arrangement.

2.1 Geometry of the System

The information about a system's geometry can usually be obtained experimentally – from an X-ray analysis of a crystal structure, NMR experiment in solution or by measuring the rotational constants in the case of small molecule complexes in the gas phase, but none of these methods is universal, and all of them contain some degree of uncertainty. For example, the position of hydrogen atoms cannot be obtained from X-ray experiments. Furthermore, it must be borne in mind that the structure of the investigated system can differ in a crystal, liquid or gaseous environment.

Optimisation to Energy Minimum. An alternative to the experiment is to obtain the geometrical parameters computationally. As the systems are usually studied in their energetical minima, geometry optimisation is one of the most common tasks in computational chemistry, during which the geometry is modified in a series of steps until the energy minimum is reached. At such a point, the forces acting on each atom are equal to zero, and the vibrational frequency analysis should yield all frequencies real and positive. Various geometry optimisation methods have been developed, the most used being the steepest descent, conjugated gradients and Newton-Raphson algorithms.

Partial Optimisation. In some cases, only partial optimisation of the system is required, in which case a part of the system is constrained while the other part is optimised. A typical case is that the positions of heavy atoms known from experiment are constrained at their initial Cartesian coordinate and the positions of hydrogens are optimised. Another frequently used possibility is to define some of the internal coordinates (bond length, angle or dihedral angle) as fixed and relax the rest of the system.

Optimisation to Transition State. The transition state (TS) structure in a chemical reaction is the structure with the highest energy (a first-order saddle point) along the reaction coordinate. Like the energy minimum, this structure is also a stationary point at the potential energy hypersurface; the gradient components are equal to zero. Unlike the energy minimum, the TS is characterised by one vibration with an imaginary frequency (one negative eigenvalue of the Hessian). This imaginary mode leads to the reactant structure in one direction and to the product in the other. The knowledge of the TS structures for various reaction mechanisms and channels makes it possible to determine the differences in TS barriers and thus to estimate the ratio of the products of the reaction performed under kinetic control.

Nevertheless, the computational determination of the TS structures has proved to be a challenging task, because it is not the energy but the gradient that has to be minimised. No method of geometry optimisation developed so far is guaranteed to converge to the TS structure. A very good supposition of the initial structure, based on the chemical intuition and trial-and-error search, is often necessary for the algorithm to attain the desired TS structure.

The recommended protocol commonly used in the TS search consists of the following steps: the reactant and product structures are first interpolated with a set of discrete values of the internal coordinate which changes the most between the reactant and the product. Second, restricted optimisations are performed on this set of structures (freezing the selected internal coordinate and optimising all the other degrees of freedom), and the initial supposition of the TS is taken as the energy maximum at this approximate reaction coordinate. Finally, the frequency analysis is calculated on this initial TS supposition structure, and optimisation to the saddle point is performed along the eigenvector corresponding to the vibration with the imaginary frequency.

2.2 Interaction Energy

The internal interaction of a molecular complex AB can be characterised by several factors. Apart from the interaction geometry (the mutual orientation of the interacting moieties A and B and their distance), another important parameter describing the interaction is its strength. The driving force for the formation of the AB complex is the Gibbs free energy ΔG of this spontaneous process, which, however, is not easily accessible computationally. Therefore, instead of the Gibbs free energy hypersurface, the potential energy surface (PES) is used. For example, the most stable geometries are found by minimising the energy (E) rather than the Gibbs free energy of the complex. According to the following equations:

$$\Delta G = \Delta H - T\Delta S, \tag{2.1}$$

$$\Delta H = \Delta E + p\Delta V, \tag{2.2}$$

the Gibbs free energy can be identified with the energy of the system in such a case where the entropy contributions (ΔS) as well as the change of the volume of the system (ΔV) can be neglected.

Interaction Energy. The interaction energy can be readily calculated as the difference between the energy of the complex and the sum of energies of the isolated systems:

$$\Delta E_{\text{int}} = E_{\text{AB}} - (E_{\text{A}} + E_{\text{B}}). \tag{2.3}$$

Although the interaction energy does not correspond to any observable quantity, it represents a useful tool for quantifying the strength of the attraction or repulsion between the two interacting systems. While the formula for interaction energy calculation is straightforward, it still presents a computational challenge, because the ΔE_{int} value is a small number resulting from the subtraction of two very large numbers. Therefore, the calculation poses large requirements on the precision of the computational method employed.

2.3 Computational Methods in Molecular Modelling

Over the years, various methods have been developed in the field of computational chemistry for the calculation of the energy and other properties of a system. These methods range from

advanced quantum chemistry methods, which are able to cover a substantial part of the electronic correlation energy, to simple empirical potentials. The aim of the following paragraphs is to outline briefly the methods available for the calculation of noncovalent interactions and discuss their suitability for the specific task of this thesis – describing the role of aromatic interactions in both biological and chemical environments.

2.3.1 Quantum Chemical Methods

The quantum chemical methods are based on solving the time-independent Schrödinger equation for the system:

$$(\hat{T} + \hat{V})\psi = E\psi \quad . \tag{2.4}$$

Using the Born-Oppenheimer approximation, the molecular wave function can be divided into the nuclear and electronic parts. With the position of the nuclei fixed, the electronic energy (which depends on the chosen position of the nuclei) can be calculated using the electronic Schrödinger equation:

$$\hat{H}_e \chi(r;R) = E_e \chi(r;R), \tag{2.5}$$

where r are all electronic coordinates and R are the (parametric) positions of the nuclei.

The Schrödinger equation of complex systems, such as atoms and molecules, cannot be solved precisely; only an approximate solution can be obtained. The quantum chemical methods can be divided into two groups according to their approach to finding the approximate solution – variational and perturbational. In the first group of methods, the variational theorem is used, which states that the energy E of any trial or approximate wave function ψ is always greater than or equal to the exact ground-state energy E_0. Besides the Hartree-Fock theory, the variational methods include the configuration interaction and coupled cluster methods. In the second group of methods, which are based on the perturbation theory, it is assumed that the exact Hamiltonian $\hat{H}$ can be divided into two parts, one that corresponds to a (known) solution of an unperturbed problem, or a "zeroth-order" Hamiltonian $\hat{H}_0$, and a small perturbation $\hat{V}$:

$$\hat{H} = \hat{H}_0 + \lambda \hat{V}, \tag{2.6}$$

where λ is an arbitrary parameter. One example of the application of the perturbation theory is the popular Møller-Plesset method. The above-mentioned methods are briefly described below.

The Hartree-Fock (HF) Method. The HF approximation is the basic level of quantum chemistry calculations. The simplicity of this method stems from its completely neglecting electronic correlation. Under the HF approximation, the system behaves as if each electron moved in the mean field created by all the other electrons (therefore it is also called one-electron approximation) and thus the wave function of the system can be approximated by a single Slater determinant. The approximation can hence be used only to determine the ground-state wave function and energy.

The HF energy for a closed-shell system has the form:

$$E_{HF} = V_{NN} + 2\sum_{i}^{n} h_{ii} + \sum_{i}^{n}\sum_{j}^{n} (2J_{ij} - K_{ij}), \tag{2.7}$$

where the terms have the following meaning: V_{NN} is the nuclear repulsion energy, h_{ii} are one-electron energies of each orbital, J_{ij} is the coulomb integral and K_{ij} is the exchange integral resulting from the Pauli principle. Unfortunately, the HF method is not suitable for calculating the interactions of aromatic systems, because a large portion of their interaction energy is caused by the correlation effects (dispersion).

The Møller-Plesset (MP) Method. MP is an *ab initio post*-HF method based on the Raleigh-Schrödinger perturbation theory. In the MP method, the perturbation $\hat{V}$, which represents the electron correlation, is defined as

$$\hat{V} \equiv \hat{H} - \hat{F} - \langle \Phi_0 | \hat{H} - \hat{F} | \Phi_0 \rangle, \tag{2.8}$$

where the normalised Slater determinant Φ_0 is the lowest eigenfunction of the Fock operator $\hat{F}$. The unperturbed (zeroth-order) Hamiltonian is then

$$\hat{H}_0 \equiv \hat{F} + \langle \Phi_0 | \hat{H} - \hat{F} | \Phi_0 \rangle. \tag{2.9}$$

Since the zeroth-order energy E_{MP0} is the Hartree-Fock energy, and the first-order MP energy E_{MP1} is obviously zero,

$$E_{MP0} \equiv E_{HF} = \langle \Phi_0 | \hat{H} | \Phi_0 \rangle, \tag{2.10}$$

$$E_{MP1} \equiv \langle \Phi_0 | \hat{V} | \Phi_0 \rangle = 0, \tag{2.11}$$

at least the second-order term is necessary to improve the HF energy, which leads to the popular MP2 method. The higher-order terms are computationally much more demanding. The MP3 does not bring much improvement over the MP2, the MP4 yields very good results but is much more expensive computationally than the MP2. Terms of higher than the fourth order are only rarely evaluated. The second-order Møller-Plesset perturbation treatment is advantageous in many cases, because it is relatively inexpensive while incorporating a large portion of correlation energy. It can therefore be utilised not only for isolated systems and H-bonded complexes but also for dispersion-bound systems. When combined with the resolution of identity technique (the RI-MP2 method), a very favourable accuracy/CPU time ratio can be achieved.

The Coupled Cluster Method. In the calculations of weakly-bound systems, especially where dispersion energy plays a major role, *e.g.* in the interactions of aromatic systems, the higher-order correlation energy terms are important and should be included. A practical route to the complete correlation energy limit is the usage of the coupled-cluster method with single and double excitations augmented by a perturbative triples correction (CCSD(T)). It has been shown that the stabilisation energies of the H-bonded and stacked model systems evaluated at the CCSD(T) level were practically identical to those calculated at the CCSDT level, where all single, double, and triple electron excitations are determined iteratively. As the CCSDT energies are close to the full configuration interaction limit, this makes CCSD(T) calculations, which are applicable also for larger biomolecular clusters, reliable.

The Symmetry-Adapted Perturbation Theory (SAPT). The SAPT is frequently the method of choice when studying intermolecular interactions, because it is capable not only of calculating interaction potentials of high accuracy but also of decomposing total interaction energy into physical components (electrostatic, induction, dispersion, and exchange terms), thus providing valuable insight into the nature of the intermolecular forces.

In the SAPT, the Hamiltonian of the dimer is decomposed into three parts as $\hat{H} = \hat{F} + \hat{W} + \hat{V}$, where $\hat{F}$ is the Fock operator, representing the sum of the Fock operators for the separate monomers; $\hat{W}$ is the intramonomer correlation operator, accounting for the intramonomer correlation effects; and $\hat{V}$ is the intermolecular interaction operator. Total interaction energy E_{int} is in the SAPT method calculated as a sum of electrostatic, exchange, induction, and dispersion components, with the dispersion and induction components also having their exchange counterparts:

$$E_{\text{int}} = E^{1}_{\text{pol}} + E^{1}_{\text{ex}} + E^{2}_{\text{ind}} + E^{2}_{\text{ex-ind}} + E^{2}_{\text{disp}} + E^{2}_{\text{ex-disp}} + \delta\text{HF}. \quad (2.12)$$

The exponents in Equation (2.12) refer to the perturbation order with respect to the intermolecular operator $\hat{V}$. δHF denotes the estimate for higher-order contributions.

It is, nevertheless, a known fact that the wave-function-based *ab initio* methods such as the conventional SAPT are computationally expensive and the calculations of larger systems are beyond the reach of this method. To reduce the cost of such calculations, the SAPT(DFT) method, based on the density functional theory (DFT) description of interacting monomers, has been introduced. In this method, it is necessary to circumvent the common failure of DFT methods to describe correctly the dispersion interaction. This drawback, which occurs due to the wrong long-range behaviour of the electron densities in commonly used exchange-correlation potentials, can be solved by an asymptotic correction to the exchange-correlation potential. Furthermore, the DFT method is only used to describe isolated monomers, and the interaction energies are calculated at a higher level. It has been shown that the corrected SAPT(DFT) can provide an accuracy of interaction energy calculation similar to high-level wave-function based methods with extrapolations to the complete basis set limit.

2.3.2 The Density Functional Theory

The Density Functional Theory (DFT) methods provide a useful alternative to the wave-function HF method. They are even less demanding computationally, because the many-body wave function of the system is not calculated; instead, the molecular properties

(*e.g.* the energy of the system) are calculated as a functional of electronic density. This leads to the Kohn-Sham equations, which, like the Hartree-Fock equations, are solved iteratively.

The energy of the electron gas can be expressed as a function of electron density:

$$E[\rho] = T[\rho] + \int V_{ext}(\vec{r})\rho(\vec{r})d\vec{r} + V_H[\rho] + E_{xc}[\rho], \tag{2.13}$$

where T is the kinetic energy of the electron gas, V_{ext} is an external potential acting on the system, V_H is the Hartree energy and E_{xc} is the exchange-correlation energy, which includes terms accounting for both exchange energy and the electron correlation. The exact exchange-correlation functionals are not known except for free-electron gas. Various DFT methods used for chemical calculations differ in which function is used for E_{xc}.

In the Local Density Approximation (LDA) methods, the exchange and correlation energies per particle, $\varepsilon_x(\rho(\vec{r}))$ and $\varepsilon_c(\rho(\vec{r}))$, of a homogeneous electron gas are calculated as a function of the density and used as an approximation for the exchange and correlation energy in the inhomogeneous system:

$$E_{xc}^{LDA}[\rho] = \int (\varepsilon_x(\rho(\vec{r})) + \varepsilon_c(\rho(\vec{r})))\rho(\vec{r})d\vec{r}\,. \tag{2.14}$$

Since the LDA approximation does not take the nonlocal effects (dispersion) into account, at least partial improvement to the LDA scheme is the Generalised Gradient Approximation (GGA), which takes into consideration the gradient of the density as an additional variable:

$$E_{xc}^{GGA}[\rho^{\alpha},\rho^{\beta}] = \int \varepsilon_{xc}(\rho^{\alpha},\rho^{\beta},\nabla\rho^{\alpha},\nabla\rho^{\beta})\rho(\vec{r})d\vec{r}\,. \tag{2.15}$$

The most popular gradient-corrected exchange-correlation functionals used now are the PBE, PW91 and BLYP functionals.

The GGA functionals are, however, still local and despite bringing a substantial improvement over the LDA scheme, they notwithstanding have problems when describing systems with long-range correlation, which are especially important in interactions of aromatic systems. The functionals containing more *semi*-local information than the GGA functionals are called *meta*-GGA functionals. They can be written as:

$$E_{xc}^{meta\text{-}GGA}[\rho^{\alpha},\rho^{\beta}] = \int \varepsilon_{xc}(\rho^{\alpha},\rho^{\beta},\nabla\rho^{\alpha},\nabla\rho^{\beta},\gamma(\vec{r}),...,\lambda(\vec{r}))\rho(\vec{r})d\vec{r}\,, \qquad (2.16)$$

where the new functions $\gamma(\vec{r}),...,\lambda(\vec{r})$ contain some *semi*-local information, such as the kinetic-energy density, higher-order density gradients or gradients of the Kohn-Sham orbitals. Yet the performance of such functionals for truly nonlocal phenomena should not be overemphasised. The most widely known *meta*-GGA functional is TPSS, which augments the GGA with terms dependent on the kinetic energy density.

Another way of improving the performance of DFT functionals is to incorporate a fixed amount, typically 20–25%, of the exact Hartree-Fock type exchange to the usual density functional exchange. Functionals constructed in this way are called hybrid exchange-correlation functionals. Although the hybrid functionals usually perform better in describing systems with important long-range interactions, there is still room for improvement. The most popular and widely used hybrid functional is B3LYP; other prevalent functionals include PBE1 and PBE0.

The DFT-D. As explained above, the performance of the DFT methods is poor in systems where London dispersion plays a major role because of its insufficient description of nonlocal effects. Nevertheless, the DFT can easily be corrected and parameterised to cover the dispersion term by adding an empirical dispersion term. In the resulting method, termed DFT-D, the interaction energy is constructed as a sum of the DFT interaction energy and a damped dispersion term of the $f(R)C_6/R^6$ form.

Since the DFT-D is parameterised using the CCSD(T)/CBS values, it provides reliable characteristics of isolated systems as well as H-bonded and dispersion-bound complexes. Moreover, the dispersion energy determined by the C_6/R^6 expression agrees surprisingly well with the dispersion contribution calculated using the SAPT method. Hence, the DFT-D interaction energy can be easily decomposed into the DFT part (roughly covering the electrostatic, induction and exchange-repulsion energies) and the dispersion part. Due to its favourable system-size/CPU-time ratio, it can be used also for extensive biomolecular systems. The cost of calculation can be further reduced by applying the RI approximation.

2.3.3 Empirical Force Fields

The molecular mechanics force field consists of the functional form, used to calculate the energy of the system (and its derivatives), and the parameter set. In the simplest form, the energy of a system is calculated as a sum of the terms of covalent (bonding, angle and dihedral angles) and noncovalent (electrostatic and van der Waals) terms:

$$E = E_{\text{bond}} + E_{\text{angle}} + E_{\text{dihedral}} + E_{\text{electrostatic}} + E_{\text{vdW}}. \quad (2.17)$$

The van der Waals term, covering the dispersion energy, is usually calculated as the Lennard-Jones 12-6 function, which has the following form for the interaction between two atoms:

$$E_{\text{vdW}}(r) = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right]. \quad (2.18)$$

The parameterisation of a force field is usually performed by fitting the parameters to both experimental results and high-level quantum calculations. The interaction energy can be calculated simply as the sum of intermolecular nonbonded terms, because all other terms in Equation (2.17) cancel each other out.

The great advantage of the force field methods is that they are several orders of magnitude faster than quantum mechanics methods while their accuracy may still be satisfactory due to parameterisation. Nevertheless, this relative reliability only holds true for the type of systems for which the parameters have been derived, and careful testing is required if the method is to be applied to other systems.

Popular force fields include Cornell *et al*. force field implemented in the AMBER package, designed mainly for proteins and DNA, the CHARMM force field, used for both small molecules and biomolecules, and GROMACS force field, which is widely used for both biological macromolecules and polymers.

2.3.4 Procedures to Increase Speed and Precision

The Counterpoise (CP) Correction. Interaction energies calculated with small basis sets can be affected by the basis set superposition error (BSSE). This error, which is a result of the

finite basis set size and vanishes asymptotically as the complete basis set is approached, arises from the fact that the wave function of the complex is expanded in more bases than the wave functions of the monomers. This leads to an improved description of the complex when compared to the monomers, yielding an artificially increased binding energy.

Despite several methods having been proposed for the elimination of the BSSE, such as the Chemical Hamiltonian Approach (CHA), which *a priori* prevents the mixing of basis sets of different monomers, the conceptually simplest method is the counterpoise method of Boys and Bernardi. In this method, the monomer energies are calculated in the basis set of the whole complex by introducing "ghost orbitals". The interaction energy can then be simply calculated as:

$$E_{\text{int}} = E_{\text{AB}}(\text{AB}) - E_{\text{A}}(\text{AB}) - E_{\text{B}}(\text{AB}), \tag{2.19}$$

where E_{AB} is the energy of the complex AB, and E_{A} and E_{B} are the energies of the monomers. The parentheses denote that a basis set of the whole complex is used in all three cases.

The CBS extrapolation. For the highly accurate calculation of interaction energy, the wave function has to be expanded as completely as possible not only in the *n*-electron space but also in the one-electron space. Therefore, in addition to using a highly precise correlation method, the complementary problem of one-electron basis set convergence should be addressed. The usage of an extended basis set is especially important for the calculation of noncovalent interactions. In order to minimise the error resulting from using a finite basis set, extrapolation to the complete one-electron basis set (CBS) limit can be performed based on the convergence behaviour. It is, however, impractical to perform the CBS extrapolation at the CCSD(T) level as the calculations for clusters of biologically relevant molecules at his level of theory would be too computationally demanding using the basis sets of the required size for extrapolation. Fortunately, the difference between CCSD(T) and MP2 energies shows very little basis set dependence (unlike these energies themselves). Therefore, one can approximate the CCSD(T)/CBS interaction energy as:

$$E_{(\text{CCSD(T)/CBS})} = E_{(\text{MP2/CBS})} + CCSD(T)_{\text{corr}}, \tag{2.20}$$

where $E_{(\text{MP2/CBS})}$ denotes the CBS limit of the interaction energy at the MP2 level and $CCSD(T)_{\text{corr}}$ is the almost basis-set-independent correction term, which can be determined as the difference between MP2 and CCSD(T) interaction energies using a smaller basis set.

Several extrapolation schemes have been developed for the determination of the MP2/CBS energy. Since the extrapolation can only be performed when systematically improved AO basis sets are used, Dunning's correlation-consistent polarised basis sets cc-pV*x*Z (*x* = D, T, Q) or their versions augmented with diffuse functions, aug-cc-pV*x*Z, are often utilised. The two-point extrapolation scheme suggested by Helgaker *et al*. is probably the most commonly used today. The described CCSD(T)/CBS procedure excludes the traditional problems of *ab initio* quantum chemical calculations, *i.e.* the incompleteness of the AO basis set and insufficient amount of correlation energy covered. The BSSE should be by definition negligible at the CBS limit.

The Frozen Core Approximation (FCA). In the *post*-HF methods, the FCA is commonly used to shorten the computational time by reducing the number of variational parameters which need to be optimised. This approximation rests on the assumption that the chemically important electrons are the valence ones and that the core electron orbitals change only negligibly upon the change of the chemical environment. Consequently, the lowest-lying molecular orbitals are constrained to remain doubly occupied in all configurations.

The Resolution of the Identity (RI) Approximation. The evaluation of the two-electron four-centre Coulomb (RI-J) integrals over Gaussian basis functions is a significant component of the overall computational time of many *ab initio* methods used nowadays. The basic approach of the RI method is to factorise the four-centre integral into three-centre quantities using a second or "auxiliary" basis set. This is formally done by inserting a resolution of identity $I = \sum_i |i\rangle\langle i|$ into the two-electron integrals:

$$\langle ij|kl\rangle \approx \sum_t \langle ij|t\rangle\langle t|kl\rangle . \tag{2.21}$$

Owing to the incompleteness of the actual auxiliary basis set, the expansion introduces an error, which should be minimised. In practice, the RI-MP2 method, for example, yields energies almost identical to the exact MP2 method with the time saving being as high as one order of magnitude. In commonly used program packages, this technique is also available for the HF and DFT calculations.

2.4 Continuum Solvent Models

Since most chemical processes take place in solvent, it is often desirable not to model the studied system isolated in its "gas phase" but to include the effects of the environment in the model. The solvent can be modelled explicitly, but this method considerably increases the computational demands of the calculation. In many cases, the solvent mainly acts as a "bulk medium", influencing the solute mainly through its dielectric properties. In such a case, it is advantageous to use an implicit, or continuum solvent model. The solvation models make it possible to evaluate the solvation free energy ΔG_{solv}, the free energy of the solute-solvent interactions, which equals to the free energy of the transfer of the solute molecule from vacuum to solvent. The ΔG_{solv} can be decomposed into several contributions:

$$\Delta G_{solv} = \Delta G_{elst} + \Delta G_{vdW} + \Delta G_{cav}\,, \tag{2.22}$$

where ΔG_{elst}, ΔG_{vdW} and ΔG_{cav} are the electrostatic, van der Waals and cavitation contributions, respectively.

The van der Waals and cavitation terms are often combined and estimated as linearly proportional to the solvent-accessible surface area of the solute, based on the experimentally determined free solvation energies. The electrostatic contribution is very important for charged and polar solutes, due to the polarisation of the solvent. The electrostatic contribution is evaluated through the solvent being modelled as a uniform medium with the dielectric constant ε. These calculations are usually based on the models derived by Born and Onsager, which describe a charge and a dipole in a spherical cavity, respectively.

In the Onsager model, the dipole of the solute induces a dipole in the surrounding medium, which in turn induces an electric field in this cavity (a reaction field). This model can be used with the quantum mechanics calculations where the interaction of the solute dipole with the reaction field is considered as a perturbation of the Hamiltonian of an isolated molecule. The drawback of this method, referred to as the self-consistent reaction field (SCRF), is the use of a spherical cavity. A more realistic cavity shape based on the van der Waals radii of the atoms of the solute is used in the polarisable continuum method (PCM). Unlike in the SCRF method, the ΔG_{elec} needs to be evaluated numerically here. The cavity surface is divided into a large number of surface elements, and a point charge representing the solvent polarisation is associated with each element. The Conductor-like Screening Model (COSMO) of a solvent,

implemented in the Turbomole program, is a variant of the PCM method, where the cavity is considered to be embedded in a conductor with an infinite dielectric constant. The potential on the surface of the conductor is set to zero, which gives rise to a convenient boundary condition for the determination of the surface charges.

In molecular mechanics calculations, the generalised Born (GB) approach is often used to evaluate the ΔG_{elec} term. Here the solute is represented as a set of particles with effective Born radii a_i and charges q_i. This model has also been incorporated in the QM calculations. The popular GB/SA model is a GB model with a cavitation and van der Waals correction based on the surface-accessible area.

Charles University in Prague
Faculty of Natural Sciences
Department of Organic and Nuclear Chemistry

Noncovalent interactions of aromatic systems and their role in proteins and organocatalysis

Doctoral Thesis

Lada Biedermannová

Supervisors:
RNDr. Zdeněk Havlas, DrSc.
RNDr. Jiří Vondrášek, CSc.

Institute of Organic Chemistry and Biochemistry of the ASCR
Center for Biomolecules and Complex Molecular Systems

Univerzita Karlova v Praze
Přírodovědecká fakulta
Katedra organické a jaderné chemie

Nekovalentní interakce aromatických systémů a jejich význam v proteinech a organokatalýze

Disertační práce

Lada Biedermannová

Školitelé:
RNDr. Zdeněk Havlas, DrSc.
RNDr. Jiří Vondrášek, CSc.

Ústav organické chemie a biochemie AV ČR
Centrum biomolekul a komplexních molekulárních systémů

I hereby declare that the doctoral thesis being presented has been elaborated solely by myself and all the literature used is properly cited. Neither the thesis nor its parts have been used before for obtaining any academic degree.

Prague, 15th January 2008 Lada Biedermannová

I would like to thank Dr Zdeněk Havlas for his interest in my research and also for his helpful comments on the text of the thesis. Many thanks are owed to Dr Jiří Vondrášek for his guidance and support in the projects involving proteins. I also acknowledge my debt of gratitude to the members of the departments of Theoretical Chemistry and Molecular Modelling, above all to Prof Pavel Hobza and to those colleagues who have collaborated with me on various projects, namely Vojtěch Klusák, Dr Petr Jurečka and Dr Kevin E. Riley. Last but not least, I am also grateful to Prof Pavel Kočovský and Dr Lubomír Rulíšek for their advice and leadership in the organocatalysis project.

Contents:

1 Preface

1.1 Subject of the Thesis

The subject of this thesis is the interactions of aromatic groups and their role in the processes of chemical and biological recognition and noncovalent binding. The importance of aromatic noncovalent interactions arises from the ubiquitous occurrence of aromatic groups. They play an important role in many processes and structures of both biological and chemical origin. In proteins, the interactions of the aromatic amino acids (phenylalanine, tyrosine, tryptophane) contribute to protein folding and protein structure stabilisation as well as to the recognition of a ligand molecule by the receptor protein. In DNA, the double-helical architecture is maintained not only by hydrogen bonds between bases but also by their stacking interactions. It has been shown recently that these interactions can also govern the self-assembly of synthetic building blocks in nanomaterials. Due to their unique properties, the aromatic substituents (phenyl, naphtyl) are indispensable also in organic reactions.

The characteristic properties of aromatic groups arise from their structure. The prerequisites of aromaticity are described by the well-known Hückel rule (first succinctly formulated not by Hückel himself but by Doering only in 1951): the group has to be a planar ring system with 4n+2 delocalised π orbitals, *i.e*. alternating single and double bonds. These requirements result in the characteristic properties of the aromatic groups – the face of the aromatic group is a relatively large flat surface characterised by a negative electrostatic potential, whereas the hydrogens on the edge bear a partial positive charge.

The described properties greatly affect the types of noncovalent interactions in which the aromatic moiety typically participates. Due to the size of the system and the large electronic density at the ring face, the dispersion interaction is dominant while the charge distribution may affect the directionality of the interaction. In a benzene dimer, for example, the two most stable structures are the parallel displaced and the T-shaped ones (see Figure 1.1), in both of which the quadrupole–quadrupole interactions are favourable. The substituents affect both the strength and directionality of the interaction, as described by the classical Hunter-Sanders rules. However, it has recently been shown that these rules derived using a simple electrostatic model are not applicable in all cases.

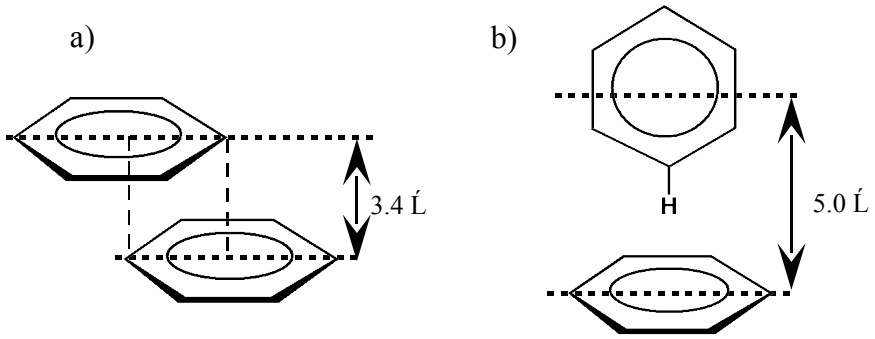


Figure 1.1: Stable arrangements of a benzene dimer: a) parallel displaced; b) T-shaped.

Both experimental and theoretical methods can be used to study aromatic interactions. While these two approaches generally complement each other, the real data in question are usually not directly comparable and great caution is needed when the techniques are used together to draw conclusions. The contradiction between these two approaches arises from the fact that whereas the experimental techniques mainly yield ensemble average values, computer simulations deal with single molecules. Hence the experimentally observable quantity is the difference in Gibbs free energy, the driving force of all chemical processes, which requires *e.g.* calorimetric studies for it to be possible to decompose into enthalpy and entropy parts. Computer simulations, on the other hand, provide direct information about the enthalpy (energy) of the studied systems, but the entropy part is computationally very difficult to obtain. Computational studies are thus useful for the investigation of the energetical part of the driving forces in chemical processes and also make it possible to model structures which cannot be observed experimentally. Last but not least, theoretical modelling is less demanding in terms of time and equipment when compared to most experimental techniques.

The aim of this work is to investigate aromatic interactions using computational methods in two specific areas. A greater part of the thesis is devoted to exploring the role of aromatic interactions in proteins, namely the role of aromatic residues and their various interactions in the stabilisation of protein structure. The second part focuses on the role of aromatic–aromatic interactions in the enantiodifferentiation in reactions catalysed by chiral organic catalysts.

1.2 Structure of the Thesis

The thesis opens with a brief overview of the computational methods available and their suitability for the study of noncovalent interactions of the aromatic moiety (Chapter 2). The

main body of the thesis is divided into several chapters, each dealing with a specific project, which generally corresponds to one of the attached research papers. The following three chapters deal with aromatic interactions in proteins from different perspectives. Chapter 3 discusses the role of aromatic interactions inside a hydrophobic core of a globular protein. In Chapter 4, this issue is further elaborated and the importance of aromatic residues in protein stabilisation is shown in the context of the interactions of all residues in a protein structure. Chapter 5 focuses on a few more specific interactions: the interaction between an aromatic ring and the π-electronic system of the peptide bond and the interaction between tryptophane and proline residues. In Chapter 6, the role of aromatic interactions in organocatalysis is investigated using the same principles as employed in the previous chapters. In the last chapter, general conclusions drawn from the entire work are summarised. The thesis is complemented by six original research papers published in or submitted to international journals on the relevant subject, which are enclosed in the Appendix.