Chapter 1

All Things Noncovalent

1.1 Introduction

Noncovalent interactions are an extremely important subset of the chemical processes occurring in the world. In biology, selectivity and recognition are achieved primarily through noncovalent contacts. In chemistry, noncovalent interactions influence virtually every chemical reaction and the design of building blocks held together by noncovalent bonds comprises an entire field, known as supramolecular chemistry.[1,2] In both chemistry and biology, the majority of work on this subject to date has focused on interactions in solution, where the results are strongly influenced by the presence of a solvent. However, more recently the study of supramolecular chemistry has taken to the gas phase.[3] Similarly, the study of biological systems is increasingly performed with the aid of gas phase techniques.[4] The results presented in this thesis expand upon the gas phase study of noncovalent interactions, and emphasis is placed on utilizing molecules and principles from supramolecular chemistry to study biological molecules.

Electrospray ionization mass spectrometry (ESI-MS) is the primary experimental method used throughout this work.[5] With this technique, molecules of low volatility can be introduced into the gas phase from a solution of water and methanol. Ions are typically produced by the addition or removal of a proton. This method of ionization is gentle, allowing for the gas phase observation of noncovalent complexes formed originally in

solution.[6] Various theoretical methods can be applied to quantitatively assess the energetics of these complexes in the gas phase. Density functional theory and semi-empirical calculations are used extensively in combination with molecular dynamics to evaluate the energetics of and determine structures for the noncovalent complexes studied in this thesis.

Background. All noncovalent interactions were *not* created equally; therefore, we will briefly review those that are most important to the present work. Coulombic interactions are very important in the gas phase because they are both strong and long range. The 1/r dependence for coulombic interactions allows them to operate over long distances in the gas phase. In solution on the other hand, solvents with high dielectric constants (particularly water) will substantially mediate the importance of coulombic interactions. Salt bridges are an excellent example of a coulombic interaction that serves to elegantly demonstrate this difference.

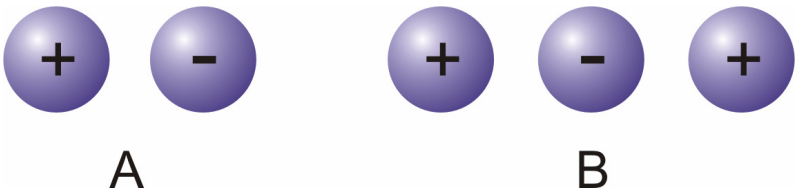

Figure 1.1 Typical salt bridges in proteins (A) and in the gas phase (B).

Minimally, a salt bridge is defined as a favorable interaction between charged functional groups (as shown generically in Figure 1.1). Biologists usually refer to the attractive interaction between (for example) protonated lysine and deprotonated aspartic acid as a salt bridge (see Figure 1.1A). By contrast, in the gas phase a salt bridge will

usually consist of three charged groups, two positive and one negative or vice versa, so that overall there is a net charge allowing the ion to be easily manipulated by electric and magnetic fields. Most of the salt bridges in the present work will be similar to that shown in Figure 1.1B, which can also be thought of as a charge stabilized ion pair.

A straightforward example is given in structure **1.1**. The binding energy of an iodide anion to the doubly charged crown ether is calculated to be ~150 kcal/mol in the gas phase (at the PM5 semi-empirical level of theory). This value is larger than most covalent bond strengths, demonstrating the strength of coulombic interactions in the absence of solvent mediation. This high binding energy is confirmed by collisional activation of **1.1**, which leads exclusively to the breaking of covalent bonds and the loss of methyliodide (see Figure 1.2). However, **1.1** is also an unusual system because there are no labile protons present that could disrupt the salt bridge. Typically in biological systems there will be labile protons present that can neutralize charges by transferring from a protonated basic site to a deprotonated acidic site. For an isolated acid/base pair, charge separation is typically not favored over proton transfer in the gas phase, but many factors can easily stabilize charge separation as explained in Chapter 3.

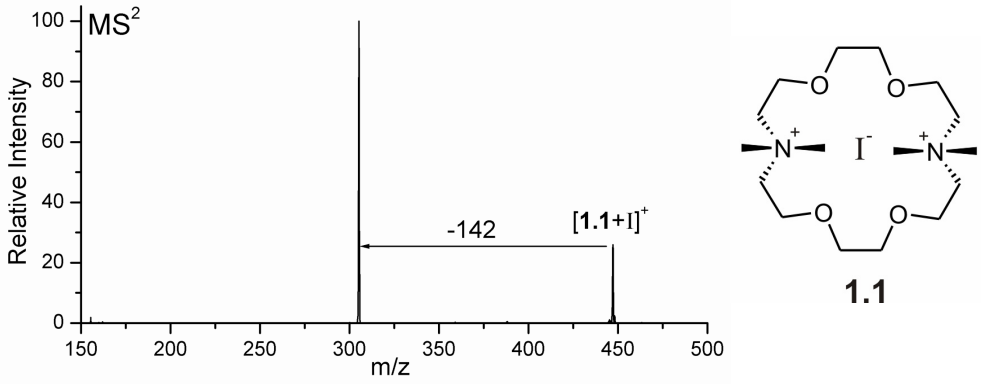

Figure 1.2 CAD of [**1.1**+I]$^+$ leads exclusively to the loss of MeI (mass 142Da).

Conversely, the stability of charge separation in solution is greatly enhanced, but the binding strength of the interaction is greatly reduced. As a result the removal of iodide anion from **1.1** in aqueous solution does not result in any covalent bond cleavage. It is for this same reason that the role salt bridges play in stabilizing protein structure in solution remains hotly debated.[7,8] Stabilization free energies vary widely, but even the highest estimates are not more than 5-10 kcal/mol for a single salt bridge in aqueous solution.[9]

Similar arguments can be made for other coulombic interactions such as ion-dipole and dipole-dipole noncovalent bonds. A special case that requires further comment is the hydrogen bond. Hydrogen bond strengths are much higher in the gas phase than in water or other polar solvents. Typical values for hydrogen bonds in water range from 2-10 kcal/mol, whereas hydrogen bond strengths up to 20 kcal/mol are possible in the gas phase.[10] Hydrogen bonds to a charged donor or acceptor can lead to hydrogen bonds up to 45 kcal/mol in the gas phase.[11] Therefore, hydrogen bonds are potentially much stronger and more important in the gas phase than in aqueous solution. Furthermore, the combination of just a few hydrogen bonds in the gas phase can easily equal the energy of a typical covalent bond (~85 kcal/mol) under the right circumstances. It should also be noted that all coulombic interactions are directional, with the strongest forces being achieved by the most linear arrangement of charges or partial charges. This directionality can be particularly important for hydrogen bonds.

Finally, another important though poorly understood solution phase interaction that is very relevant to the observation of ions by electrospray ionization mass spectrometry (ESI-MS) is the hydrophobic effect.[12,13] This is the driving force behind the aggregation of nonpolar molecules in aqueous solution, which is thought to be very important in

protein folding.[14] The strength of this interaction is difficult to define and certainly weaker in the gas phase than it is in solution. This is primarily due to the fact that, formally, there is no "hydrophobic" effect when there is no water. Only weakly binding Van der Waals forces remain after desolvation and introduction into the gas phase. Therefore, complexes that are held together in solution through largely hydrophobic interactions are not likely to be observed in the gas phase by ESI or matrix assisted laser desorption ionization (MALDI) experiments. However, it should be mentioned that the higher the relative hydrophobicity of a charged molecule, the more abundant it will appear in an ESI mass spectrum. This interesting phenomenon will be explained in greater detail in Chapter 7.

1.2 Content of Thesis

Clusters of Biomolecules. Chapters 2-5 contain work on small clusters of biologically relevant molecules. In Chapter 2, it is shown that the unusual properties of arginine lead to extensive noncovalent clustering of this amino acid, when sampled by ESI-MS as shown in Figure 1.3. The clusters can be formed as cations or anions, with a variety of different molecules serving as the charge carrier. Of particular interest are a series of anionic trimers which demonstrate unusual abundance (see Figure 1.4). A structure (**1.2**) in which each arginine interacts in a head-to-tail arrangement while maintaining an intramolecular bond is proposed to explain the unusual abundance of these clusters. Each arginine in the trimer is in the zwitterionic form. The stability of the zwitterionic form of arginine for clusters without a net charge is addressed further by theoretical methods in Chapter 3.

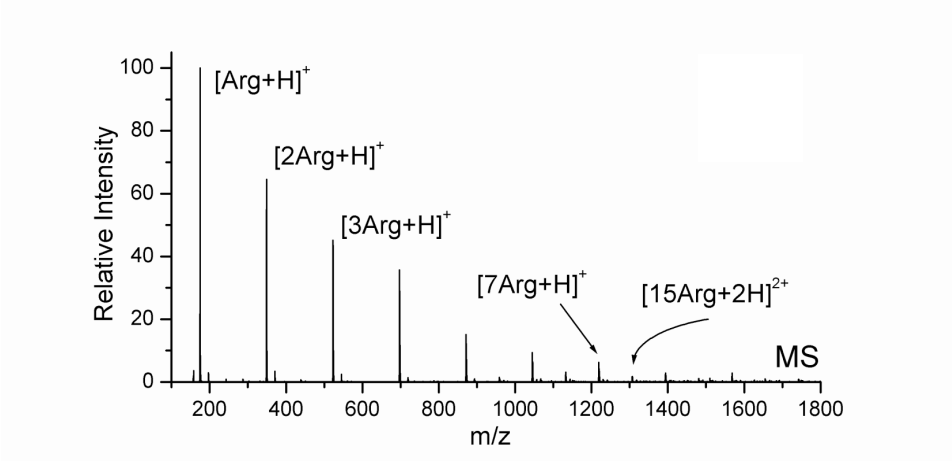

Figure 1.3 Cationic clusters of arginine.

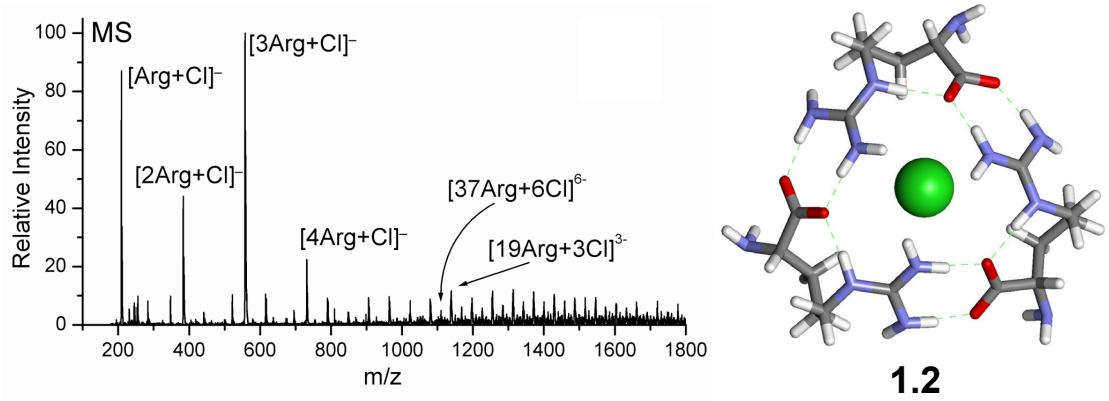

Figure 1.4 Anionic clusters of arginine with chloride.

This work on arginine clusters allows us to draw several important conclusions. The significance of the strength and specificity that can be obtained by the interaction between alkyl-guanidiniums and carboxylates is clearly demonstrated. These interactions have been observed in many crystal structures.[15,16,17] The cluster work presented in Chapter 2 establishes the importance of these interactions in the gas phase as well. There are several implications for gas phase protein structure as a result of this observation. Salt bridges between arginine and either aspartic or glutamic acid are likely to remain as charge separated salt bridges in the gas phase. Furthermore, this class of salt bridge is

predicted to be more specific and strongly bound in the gas phase than salt bridges involving lysine residues. These factors should be taken into account when molecular modeling is utilized to examine the gas phase structures of proteins.

In Chapter 4, the properties of another unusually abundant cluster are examined. The ESI-MS spectrum of a 0.01M solution of serine reveals an unusually abundant protonated octamer. Further experiments employing isotopic labeling demonstrate that this octamer has a strong preference to be homochiral. These startling observations have attracted the attention of several groups,[18] each with different structures and explanations for the observed characteristics of the serine octamer. Utilizing the hierarchy of interactions outlined in the introduction above, we constructed a serine octamer that maximized favorable coulombic interactions and hydrogen bonds. The resulting structure is cubic and has a zwitterionic core as shown in Figure 1.5. A recent review on the subject critically compared the energetics of the structures from each group and found ours to be the lowest energy conformation by a significant margin.[19]

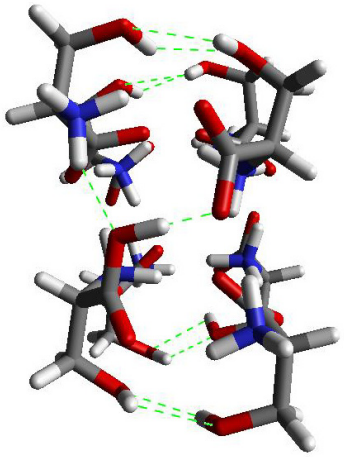

Figure 1.5 Structure for the homochiral protonated serine octamer.

The results gathered from the serine octamer demonstrate that a homochiral preference can exist for very small clusters or "nanocrystals." Furthermore, ESI may offer a new

experimental technique for investigating the early stages of homogeneous crystal nucleation in solution. A universal theory for the explanation of homogeneous nucleation is still lacking.[20] Several spectroscopic methods have been employed to study crystal nucleation,[21] but mass spectrometry offers the additional ability to sample small clusters and study them in the gas phase. Moreover, methods for symmetry breaking in racemic mixtures are very important to understanding the origin of life.[22] The work in Chapter 4 clearly demonstrates that symmetry breaking can be achieved in small molecular clusters. If some means of activation that leads to polymerization of the cluster components can be achieved, then a route to the generation of homochiral polymers is possible. While this may sound rather outlandish, polymerization reactions for biological molecules from small molecular clusters are reported in Chapter 5.

The first gas phase synthesis for ATP (adenosine triphosphate) is given in Chapter 5. This extremely important molecule is easily synthesized in the gas phase from a cluster of three AMP (adenosine monophosphate) molecules bound together by a sodium salt bridge. Subsequent collision activated dissociation (CAD) spectra following the gas phase synthesis are identical to those obtained from an authentic sample of ATP in separate experiments. It is further shown that similar chemistry is possible with phosphate itself, allowing for the gas phase generation of polyphosphate (which is another biologically important molecule).[23]

Natural processes such as sea spray could theoretically lead to the occurrence of such desolvated clusters in the Earth's atmosphere.[24] The sun provides an ample supply of energy, which could initiate the chemical reactions that lead to polymerization. Given that ATP can be synthesized by such a process and that the serine octamer provides

evidence for symmetry breaking, the chemistry of small clusters of biological molecules and their precursors probably deserves more attention.

Highly Reactive Chemistry in Clusters. Chapter 6 studies the fundamental properties of the chemistry of carbenes with emphasis on the important Wolff rearrangement. Metal ion coordination is utilized to facilitate both the generation of carbenes from diazo malonate precursors and the subsequent multiple Wolff rearrangements that follow. Isotopic labeling is employed to determine the mechanisms for the various reactions and confirm that rearrangement does not proceed through an oxirene intermediate. The influences that conformation and metal ion coordination have on the Wolff rearrangement are studied experimentally and theoretically. Reactions between these carbene species and various noncovalent adducts are also examined. The end result is a more detailed understanding on the fundamental aspects of a process which is very important to synthetic organic chemistry.

Molecular Recognition in Biological Systems. Chapters 7-9 deal with the molecular recognition of amino acid side chains through noncovalent attachment in ESI-MS experiments. In Chapter 7, the ability of 18-crown-6 ether (18C6) to recognize and selectively attach to lysine residues is explored. It is found that the number of lysines can be quantified for small peptides. For proteins, the number of 18C6 ethers that attach is related to the structure of the protein in solution, with more crowns attaching to unfolded proteins. Furthermore, 18C6 is shown to enhance the ESI signal for the ion to which it attaches by effectively desolvating the charge and increasing the surface activity of the ion on the highly charged electrospray droplet. This leads to the observation of $[KKKK+4(18C6)+4H]^{4+}$ as the base peak in the spectrum from a solution containing a

1:1 mixture of KKKK and 18C6. This structure is shown in Figure 1.6. Clearly, the solution phase compositions are not accurately reflected in the ESI-MS data for these systems due to preferential sampling of ions that are coordinated to 18C6.

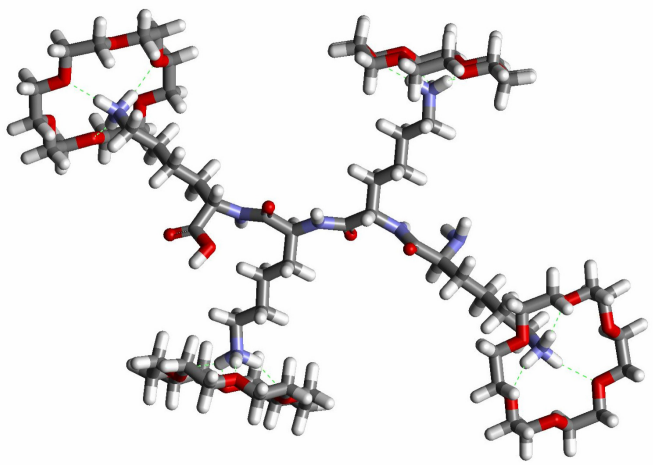

Figure 1.6 Tetralysine with 4 18C6 ethers attached.

In Chapter 8, recognition of arginine side chains is accomplished in a similar manner by utilizing the larger dibenzo-30-crown-10 ether (DB30C10) as shown in Figure 1.7. This crown preferentially recognizes the side chain of arginine, but does not form highly abundant adducts like 18C6 does with lysine. The reason for this is unclear, but the net result is that only one arginine will be reliably identified on a peptide that may contain multiple arginines. It is shown through competitive CAD experiments that the larger crown has a higher binding energy to arginine than 18C6 does to lysine. Furthermore, the techniques in Chapter 7 and 8 are mutually compatible, allowing for both crowns to be added to the same solution. This should allow for the easy separation of a tryptic digest into the lysine and arginine containing fragments without *a priori* knowledge of the sequences of the peptides. Such a technique may be useful in confirming the identity of a

peptide which had undergone post-translational modification and appeared at a mass other than the expected value.

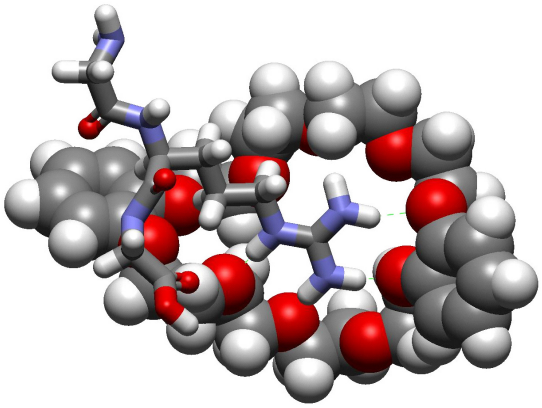

Figure 1.7 Interaction between dibenzo-30-crown-10 and the peptide GRG.

In Chapter 9, the most effective method for adding lariats to 18C6 without disrupting its excellent recognition abilities are systematically explored for reasons that will become obvious in Chapters 10 and 11. The simplest way to obtain a lariat crown ether is to begin with aza-18C6 and attach a functional group to the nitrogen heteroatom. Unfortunately, the high proton affinity of the secondary amine destroys the recognition ability of crown for gas phase experiments. It is shown that conversion of the amine to an amide reduces the effect, but the overall binding energy is still substantially lower when compared to 18C6. Interestingly, aza-18C6 is shown to have a proton affinity ~10 kcal/mol higher than any other secondary amine in the NIST database.[25] This unusually high proton affinity is due to intramolecular hydrogen bonds within the crown. With regards to the addition of lariat side chains, the results indicate that the side chain must branch off of one of the carbons in 18C6. This hypothesis was confirmed in later experiments as shown below.

De Novo Biomimetic Reagents. Chapters 10 and 11 combine the recognition of 18C6 with various chemical functionalities in order to mediate peptide chemistry in the gas phase. In Chapter 10, a new class of molecules termed "molecular mousetraps" is described (see Figure 1.8). The mousetraps combine the recognition of 18C6 with the chemical reactivity of diazo groups. The resulting molecules are capable of noncovalently attaching to any molecule that contains a protonated primary amine. In the gas phase, CAD can then be utilized to activate the complex, which results in the formation of a highly reactive carbene which then preferentially inserts intermolecularly. The noncovalent complex is transformed into a covalently bound molecule by this process. Importantly, this is an example of a system where collisional activation of a noncovalent complex results in a chemical reaction rather than simple dissociation, which is typically the dominant process in the vast majority of systems.

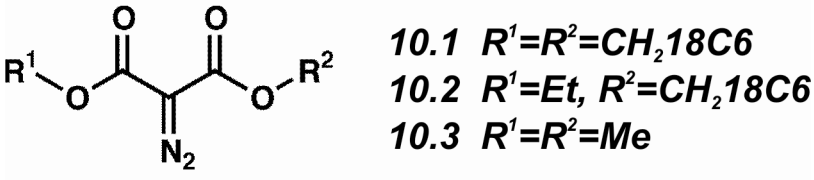

Figure 1.8 Structures of the "molecular mousetraps" in Chapter 10.

In Chapter 11, the mousetraps are utilized in experiments with peptides. It is shown that covalent attachment can be achieved in a quantitative fashion. In addition, the results for two other reagents designed to initiate peptide backbone cleavage are given. Although a directed cleavage process was not observed, progress towards the design of such a reagent was achieved. Primarily, these experiments revealed that any successful reagent must be designed with the proper combination of high binding energy and highly reactive

chemical functionalities. High binding energy can be achieved by using two 18C6 ethers to attach to a peptide containing two lysine residues. Chemical reactions with activation barriers similar to the reaction barrier for converting a diazo into a carbene appear to be optimal. These complexes are stable until triggered by CAD. If the activation energy is too low, then the process cannot be controlled and may occur prior to detection in the mass spectrometer. These initial experiments demonstrate that the de novo design of reagents capable of mediating peptide chemistry in the gas phase is possible and worth exploring further.

1.3 Summary

In conclusion, the study of noncovalent interactions can reveal fundamental information about the chemistry that is happening in the world around us. Noncovalent complexation can reveal properties of biological molecules and offers a glimpse into the possible origin of such molecules. The combination of molecular recognition with additional chemical functionalities is a promising area for the development of de novo reagents that operate in the gas phase.

1.4 References

[1] *Comprehensive Supramolecular Chemistry*, vol. 1 (Ed.:G. W. Gokel), Pergamon/Elsevier: Oxford, 1996.

[2] *Supramolecular Chemistry* Beer, P. D.; Gale, P. A.; Smith, D. K. Oxford University Press, New York, 1999.

[3] Schalley, C. A. *Int. J. Mass Spec.* **2000,** *194*, 11-39.

[4] Hoaglund-Hyzer, C.S.; Counterman, A.E.; Clemmer, D.E. *Chem. Rev*. **1999**, *99*, 3037-3079.

[5] Fenn, J.;Rosell, J.; Nohmi, T.; Shen, S.; Banks, F. *Biochemical and Biotechnological Applications of Electrospray Ionization Mass Spectrometry* **1996**, *619*, 60-80.

[6] (a) Veenstra, T. D. *Biophys. Chem*. **1999**, *79*, 63-79. (b) Loo, J. A. *Int. J. Mass Spectrom*. **2000**, *200*, 175-186.

[7] Kumar, S.; Nussinov, R. *J. Mol. Biol.* **1999**, *293*, 1241-1255.

[8] Hendsch, Z. S.; Tidor, B. *Protein Sci*. **1994**, *3*, 211-226.

[9] Lebbink, J. H. G.; Knapp, S.; van der Oost, J.; Rice, D.; Ladenstein, R.; de Vos, W. M. *J. Mol. Biol.* **1998**, *280*, 287-296.

[10] Prins, L. J.; Reinhoudt, D. N.; Timmerman, P. *Angew. Chem. Int. Ed*. **2001**, *40*, 2382-2426.

[11] *Hydrogen Bonding: A Theoretical Perspective* (Ed. S. Scheiner). Oxford University Press, New York, **1997**.

[12] Marmur, A. *J. Am. Chem. Soc*. **2000**, *122*, 2120-2121.

[13] Silverstein, K. A. T.; Haymet, A. D. J.; Dill, K. A. *J. Am. Chem. Soc*. **1998**, *120*, 3166-3175.

[14] (a) Baldwin, R.L. *Science* **2002**, *295*, 1657-1658. (b) Fernandez, A.; Kardos, J.; Goto, Y. *FEBS Letters* **2003**, *536 (1-3)*, 187-192 and references therein.

[15] Karle, J.; Karle, I.L. *Acta Cryst.* **1964,** *17*, 835-841.

[16] Bhat, T.N.; Vijayan, M. *Acta Cryst.* **1977**, *B33*, 1754-1759.

[17] Chapo, C. J.; Paul, J. B.; Provencal, R. A.; Roth, K.; Saykally, R. J. *J. Am. Chem. Soc.* **1998,** *120,* 12956-12957.

[18] The full story is detailed in Chapter 4, including the contributions of other investigators.

[19] Schalley, C.A.; Weis, P. *Int. J. Mass Spectrom.* **2002**, *221*, 9-19.

[20] Granasy, L.; Igloi, F. *J. Chem. Phys*. **1997,** *107*, 3634-3644.

[21] (a) Granasy, L.; James, P. F. *J. Chem. Phys*. **2000,** *113,* 9810-9821, (b) Peng, X.; Wickham, J.; Alivisatos, A. P. *J. Am. Chem. Soc.* **1998**, *120*, 5343-5344.

[22] (a) Podlech, J. Cell. Mol. Life Sci. **2001**, 58, 44-60. (b) Editor, David B. Cline *Physical origin of homochirality in life : Santa Monica, California, February 1995* Woodbury, New York : American Institute of Physics, **1996**.

[23] Kornberg, A.; Rao, N. N.; Ault-Riche, D. *Annu. Rev. Biochem*. **1999**, *68*, 89-125.

[24] (a) Ellison, G.B.; Tuck, A.F.; Vaida, V. , *J.Geophys. Res.* **1999**, *104*, 11633-11642. (b) Tuck, A. *Surv. Geophys*. **2002**, *23*, 379-409. (c) Gill, P.S.; Graedel, T.E.; Weschler, C.G. *Rev. Geophys.* **1983**, *21*, 903-920.

[25] E.P. Hunter and S.G. Lias, "Proton Affinity Evaluation" in NIST Chemistry WebBook, NIST Standard Reference Database Number 69, Eds. W.G. Mallard and P.J. Linstrom, February 2000, National Institute of Standards and Technology, Gaithersburg MD, 20899 (http://webbook.nist.gov)