CMLS, Cell. Mol. Life Sci. 60 (2003) 2526–2531
1420-682X/03/112526-06
DOI 10.1007/s00018-003-3280-8

CMLS Cellular and Molecular Life Sciences

Research Article

Conformational constraints on side chains in protein residues increase their information content

A. J. Bojarski[a], M. Nowak[a] and B. Testa[b,*]

[a] Department of Medicinal Chemistry, Institute of Pharmacology of the Polish Academy of Sciences, 12 Smetna St., 31343 Krakow (Poland)
[b] Institute of Medicinal Chemistry, BEP, University of Lausanne, 1015 Lausanne (Switzerland),
Fax: +41 21 692 4525, e-mail: Bernard.Testa@ict.unil.ch

Received 13 July 2003; received after revision 18 August 2003; accepted 4 September 2003

Abstract. Like all other complex biological systems, proteins exhibit properties not seen in free amino acids (i.e., emergent properties). The present investigation arose from the deduction that proteins should offer a good model to approach the reverse phenomenon, namely top-down constraints experienced by protein residues compared to free amino acids. The crystalline structure of profilin Ib, a contractile protein of *Acanthamoeba castellanii*, was chosen as the object of study and submitted to 2-ns molecular dynamics simulation. The results revealed strong conformational constraints on the side chain of residues compared to the respective free amino acids. A Shannon entropy (SE) analysis of the conformational behavior of the side chains showed in most cases a strong decrease in the SE of the χ_1 and χ_2 dihedral angles compared to free amino acids. This is equivalent to stating that conformational constraints on the side chain of residues increase their information content and hence recognition specificity compared to free amino acids. In other words, the vastly increased information content of a protein relative to its free monomers is embedded not only in the tertiary structure of the backbone, but also in the conformational behavior of the side chains. The postulated implication is that both backbone and side chains, by virtue of being conformationally constrained, contribute to the recognition specificity of the protein toward other macromolecules and ligands.

Key words. Profilin Ib; molecular dynamics simulations; conformational constraints; information; Shannon entropy; molecular recognition.

Interest in the study of complex biological systems continues to increase [1–3]. Such biological systems extend over a vast range of levels of complexity, from proteins and functional multiprotein assemblies up to organisms and societies. Yet despite this great disparity, at least two unifying characteristics are always found in complex biological systems, namely (i) the fact that they are aggregates composed of interacting (transacting) parts, and (ii) the existence in such systems of properties that do not exist in their isolated components. Such properties are termed 'emergent,' and indeed emergence is an important subject in the study of complexity [4–7].

Being relatively simple compared to biological systems at higher levels of complexity, proteins appear to be worthwhile objects to investigate the phenomenon of emergence. Like their isolated components (i.e., amino acids), proteins possess physicochemical properties such as acidity, basicity, conformational flexibility, and tautomerism, all of which allow the compounds to exist in a number of different molecular states, e.g., different electrical states and conformations [8–10]. In addition and more importantly, proteins may display biochemical and

* Corresponding author.

functional properties not seen in free amino acids, e.g., binding capacity, allosterism, and catalytic activity [11, 12].

Besides emergence, a few scientists have discussed the reverse concept, namely the constraints experienced by the components of a complex system as a result of their transactions [7, 10, 13, 14]. This phenomenon is termed downward causation, top-down constraints or dissolvence, but its occurrence has never been investigated at an explicit and quantitative level. Constraints imply decreased disorder and hence decreased entropy. And since information is a measure of order [15, 16], constraints imply an increase in information content, in other words an increase in the capacity to select from alternatives [6]. We reasoned that a good model to approach this problem would be a comparison of the behavior of side chains of amino acids in the free state and as residues in proteins. Indeed, the conformational behavior of side chains (the so-called 'Chi space') is an important determinant of the surface and recognition properties of proteins [17]. Preliminary studies using molecular dynamics (MD) simulations and analysis of three-dimensional (3D) molecular fields had shown that the property space of the central residue in tripeptides is indeed constrained compared to the free amino acid [18, 19].

Here, we present an MD simulation of profilin Ib, a contractile protein of *Acanthamoeba castellanii*. This protein contains 125 residues (reference PDB-1ACF in the Protein Data Bank). A comparison of the conformational behavior of the side chains in the residues and in the corresponding free amino acids revealed strong contraints. These constraints on conformational space were analyzed for their Shannon entropy (SE) content [20], which for most residues showed an intriguing reduction relative to the free amino acids.

Materials and methods

The coordinates of *A. castellanii* profilin Ib (contractile protein) [21] were obtained from the Protein Data Bank [22] (entry: 1ACF; resolution 2.0 Å; R factor 17.9%). This protein was chosen because of its relatively small size, suitable quality of crystallographic data, highly ordered secondary structure, and similar content of α helices and β strands (32.0% α helices, 32.8% β strands and 20.0% β turns). Furthermore, preliminary tests revealed the stability of the secondary structure during prolonged MD calculations.

The initial 3D structures of the free zwitterionic amino acids were built using Accelrys Isight II software [23]. The protonation of side chains and terminal groups corresponds to neutral pH. All MD simulations were performed using the GROMACS 3.0 [24, 25] suite of programs with a GROMACS2 all-hydrogen force field [25]. The protein and free amino acids were placed in a rectangular box of water (SPC model [26]), with the distance between box wall and solute surface not less than 1 nm. All simulation systems were subjected to (i) steepest-descent energy minimization until a gradient value of 100 kJ mol^{-1} nm^{-1}, followed by (ii) 250 ps of equilibration MD at 300 K (not included in the analysis), followed by (iii) 2 ns of production MD simulations at 300 K. The temperature and the pressure of solute were coupled to the external bath according to Berendsen's algorithm [27] with time constants 0.1 and 0.5, respectively. Time steps of 1 fs were used and no bond was constrained. Periodic boundary conditions were applied; 1.3-nm cut-off values were used to calculate Coulombic and Van der Waals interactions. The neighbor list was updated every tenth step. The coordinates of the system were collected every 0.5 ps.

The values of χ_i values were distributed in bins of 10°, and SE was calculated from these histograms using the formula

$$SE = -\sum_i p_i \log_2 p_i$$

where p_i is the probability in each bin, obtained from the count in each bin (c_i):

$$p_i = \frac{c_i}{\sum_i c_i}$$

Results

A meaningful comparison of the conformational behavior of amino acid side chains implies a constant secondary and tertiary structure of the backbone. This was indeed verified as shown in fig. 1, where the backbone of profilin Ib is seen to retain a constant secondary structure during the 2-ns simulation, and to undergo only limited oscillations in its tertiary structure.

The dihedral angles of all residue side chains in profilin Ib were recorded each picosecond, yielding a 2-ns trajectory for each dihedral angle χ of each residue side chain. The same was done for the side chain of the correspond-

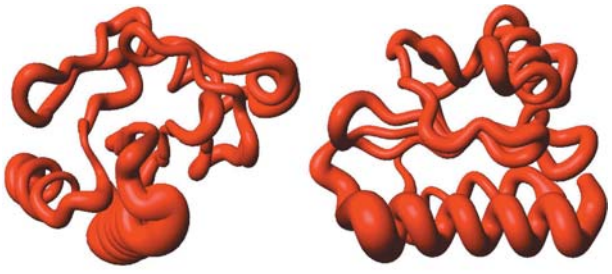

Figure 1. MOLMOL [28] orthogonal views of the 3D structure of the backbone of profilin Ib, with the backbone thickness proportional to the amplitude of its movements during the 2-ns simulation.

ing free amino acids in zwitterionic form. A representative example is presented in fig. 2, where the trajectories of χ_1 and χ_2 of the four phenylalanine residues in profilin Ib (Phe35, Phe49, Phe60, and Phe125) are compared to those of free phenylalanine. The differences are most impressive and quite revealing. Thus, the side chain of free phenylalanine shows preferred conformations characterized by $\chi_1 = -40°$ to $-80°$ and $-140°$ to $-180°$, and $\chi_2 = 60°$ to $120°$ and $-60°$ to $-120°$. The corresponding values for Phe35 are $\chi_1 = 160°$ to $180°$ and $-160°$ to $-180°$, and $\chi_2 = 30°$ to $90°$; for Phe49, $\chi_1 = \chi_2 = -40°$ to $-80°$; for Phe60, $\chi_1 = 40°$ to $80°$, and $\chi_2 = -60°$ to $-100°$; for Phe125, $\chi_1 = 30°$ to $60°$ and $-40°$ to $-90°$, and $\chi_2 = -40°$ to $-120°$. Constraints of comparable magnitude were seen for most side chains (results not shown).

The occurrence of strong conformational constraints in the side chains of profilin Ib having been demonstrated unambiguously, the next step was to quantify and interpret them. This was achieved by an SE analysis of the conformational behavior of the side chains. SE is a measure of the relative information content of a dataset, and it allows the information content of two datasets to be compared [20]. Thus, SE(1) > SE(2) means that dataset 2 contains more information than dataset 1. Shannon's theory was originally one of digital communication [29], but it is beginning to find applications in chemistry and particularly molecular-diversity analysis [30].

The values of χ_1, χ_2, χ_i of side chains (except methyl rotors) were distributed in bins of 10°, as exemplified in fig. 3 for free phenylalanine and Phe35. This allowed the SE[χ_i] to be calculated for each dihedral angle in the free amino acids and in the residues of profilin Ib. For Phe35, the difference ΔSE[χ_i] (i.e., SE[χ_i] in the residue minus SE[χ_i] in the free amino acid) was thus calculated to be –1.45 and –1.26 for ΔSE[χ_1] and ΔSE[χ_2], respectively (fig. 3). In other words, the strong conformational constraints experienced by the side chain of Phe35 produce a marked decrease in SE. The results for all side chains are presented in fig. 4 as a ΔSE[χ_1] versus ΔSE[χ_2] plot. Of the 88 χ_1 and 76 χ_2 (valines excluded) dihedral angles, the majority are located in the lower left (negative-negative) quadrant, with ΔSE[χ_1] and ΔSE[χ_2] values ranging from 0 to –2.3.

Only 6 dihedral angles had clearly positive ΔSE[χ_1] values (i.e., ΔSE[χ_1] > 0.2), and only 11 had clearly positive ΔSE[χ_2] values. Only six residues occupy the upper-right quadrant, namely Arg56, Asn111, Ser76, Ser83, Ser84, and Ser92, with ΔSE[χ_i] values between 0 and +1.0. In a free serine, the hydroxy group is H-bonded to the protonated amino group (results not shown), a conformational constraint that is partly removed in serine residues. In contrast, Ser1 retains this constraint and its SE is decreased (ΔSE[χ_1] = –0.11, ΔSE[χ_2] = –0.656). As for Ser32, its side chain points toward the protein interior and

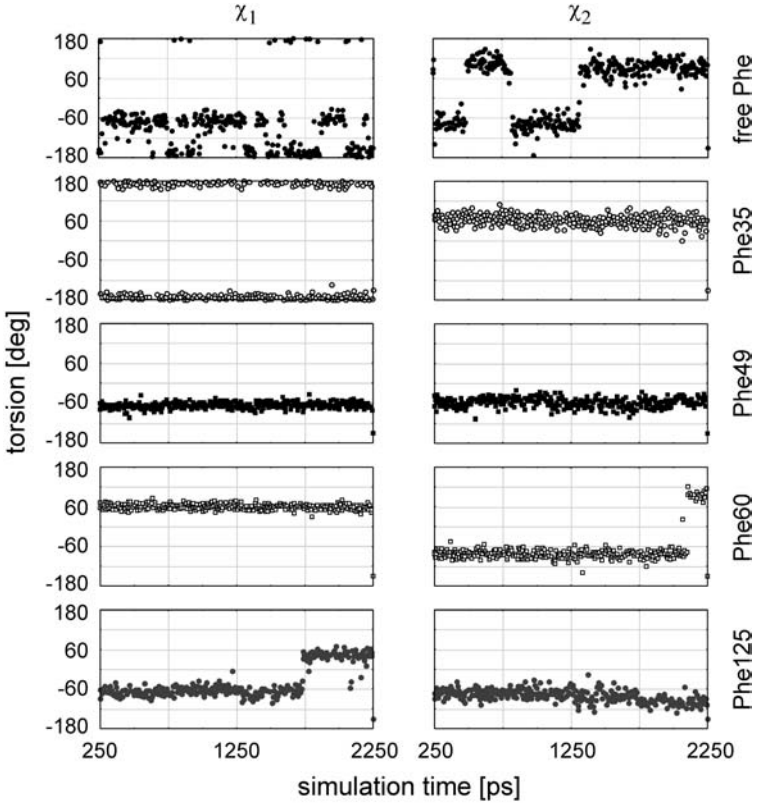

Figure 2. Trajectories of χ_1 and χ_2 in the side chain of free phenylalanine, Phe35, Phe49, Phe60 and Phe125.

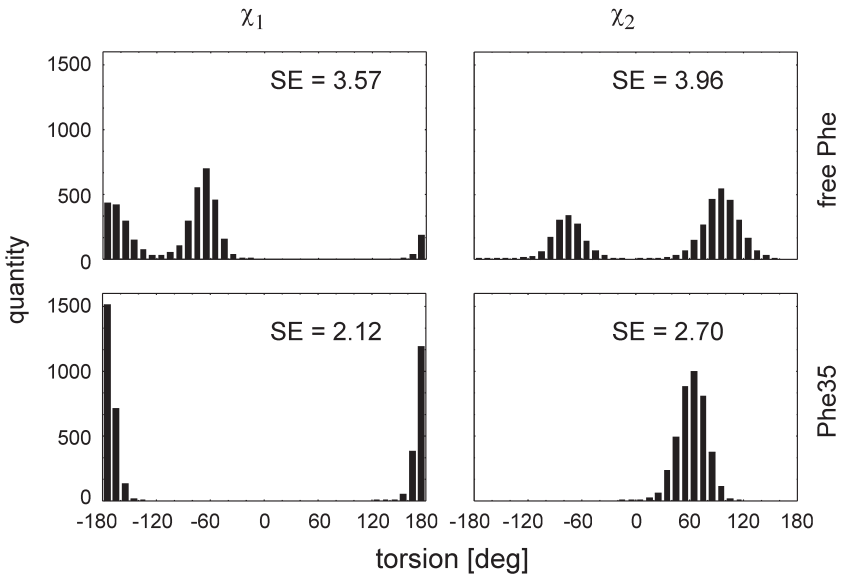

Figure 3. Calculation of SE values, as illustrated here for the dihedral angles χ_1 and χ_2 in free phenylalanine and Phe35.

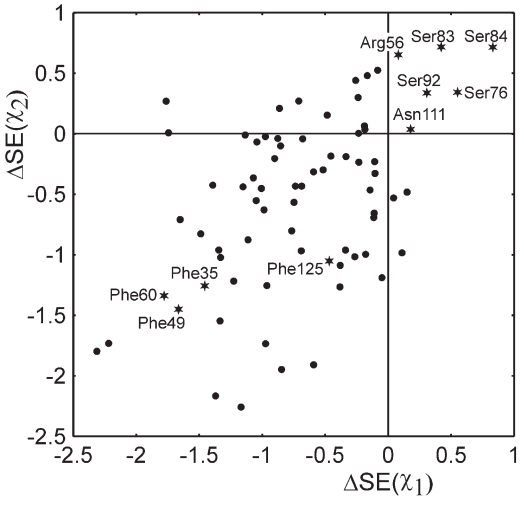

Figure 4. Plot of $\Delta SE[\chi_1]$ versus $\Delta SE[\chi_2]$, where $\Delta SE[\chi_i] = SE[\chi_i]$ in the residue minus $SE[\chi_i]$ in the free amino acid. A majority of residues are located in the lower left quadrant, with $\Delta SE[\chi_1]$ and $\Delta SE[\chi_2]$ values ranging from 0 to –2.3. In other words, residues in profilin Ib generally have a markedly lower SE content in their side chain than the corresponding amino acids in the free state.

is strongly constrained ($\Delta SE[\chi_1] = -1.227$, $\Delta SE[\chi_2] = -1.121$).

The possibility of a conformational artifact caused by comparing zwitterionic amino acids with residues (i.e., replacing charged $-COO^-$ and $-NH_3^+$ groups with –CO– and –NH–, respectively) was also examined by using (i) end-capped amino acids, and (ii) amino acids located at the center of the nonapeptide AAAAXAAAA maintained in an α-helical structure. The differences for amino acids with ionized side chains were modest or minute. Thus, the average $\Delta SE[\chi_1, \chi_2]$ value for all residues in profilin was –0.568 when using zwitterionic amino acids as reference, –0.513 when using capped amino acids, and –0.417 when comparing with the residues in the α-helical nonapeptide. The full results will be presented in a forthcoming paper. In contrast to χ_1 and χ_2, the results for the few residues with χ_3 and χ_4 dihedral angles were unremarkable, the $\Delta SE[\chi_i]$ values being modest and without pattern (results not shown).

Discussion

The MD simulations reported here demonstrate and quantify the strong conformational constraints felt by residue side chains in a protein, when compared to the corresponding amino acids in the free state. In the protein profilin Ib investigated here, these constraints involve mainly the first two dihedral angles (χ_1 and χ_2) of the side chains, with the unpredictable and remarkable consequence that the information content of these side chains, as assessed by SE, is markedly affected when comparing residues and free amino acids. In most cases, an important decrease in SE was seen. Given the inverse proportionality between SE and information, this is equivalent to stating that most side chains in residues showed an increase in information content compared to the free amino acids.

These findings raise a number of questions and suggest a broad research agenda. The deeper question is certainly the relationship, if any, between the emergence of protein-specific properties and the gain in information resulting from conformational constraints. This may not be an unrealistic hypothesis given the role of information exchange in complex biosystems [16]. But what are the

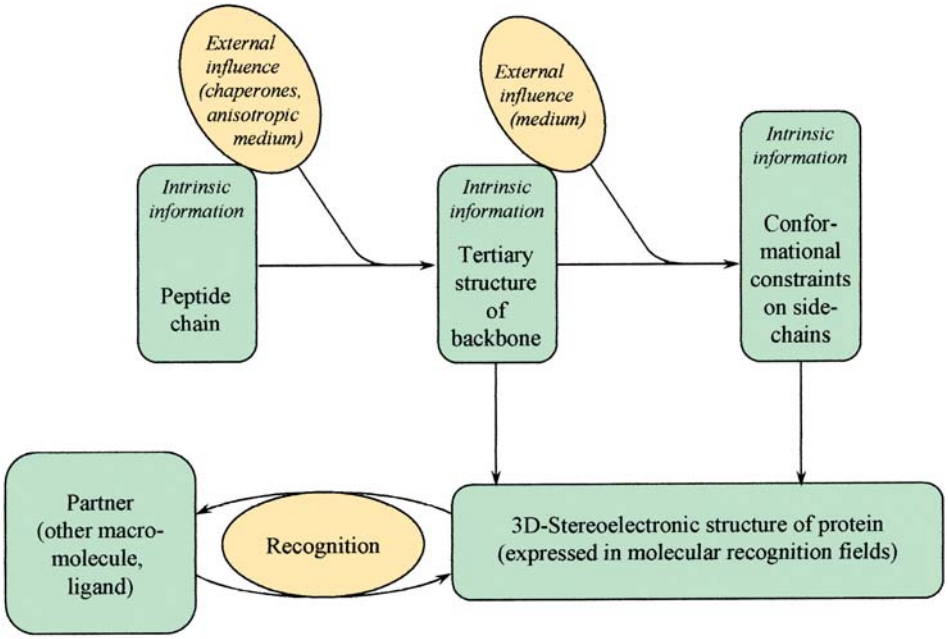

Figure 5. A schematic representation of the recognition specificity of a protein arising from the information content in its backbone and side chains.

practical implications of the interrelated conformational constraints and information content? We postulate that the increased information content of the side chain of residues compared to free amino acids contributes to the recognition specificity of the protein. In other words, the vastly increased information content of a protein relative to its free monomers is embedded not only in the tertiary structure of its backbone, but also in its side chains. The implication is that both backbone and side chains, by virtue of being conformationally constrained, contribute to protein recognition specificity toward other macromolecules and ligands (fig. 5), as cogently advocated by Loewenstein [15].

A more immediate issue is to generalize these results in a variety of structurally and functionally diverse proteins. Preliminary results with other functional proteins confirm the general nature of the phenomenon reported here, but experiments on a larger scale are needed to evaluate its significance and meaning along the lines outlined in the previous paragraph.

1 Chong L. and Ray L. B. (2002) Whole-istic biology. Science **295:** 1661
2 Kitano H. (2002) Systems biology: a brief overview. Science **295:** 1662–1664
3 Csete M. E. and Doyle J. C. (2002) Reverse engineering of biological complexity. Science **295:** 1664–1669
4 Prigogine I. and Stengers I. (1985) Order out of Chaos, Fontana, London
5 Crutchfield J. P. (1994) Is anything ever new? Considering emergence. In: Complexity: Metaphors, Models, and Reality, pp. 515–533, Cowan G., Pines D. and Metzler D. (eds), Addison-Wesley, Reading, Mass.
6 Gell-Mann M. (1995) The Quark and the Jaguar, Abacus, London
7 Luisi P. L. (2002) Emergence in chemistry: chemistry as the embodiment of emergence. Found. Chem. **4:** 183–200
8 Platt J. R. (1961) Properties of large molecules that go beyond the properties of their chemical sub-groups. J. Theor. Biol. **1:** 342–358
9 Prigogine I. (1978) Zeit, Struktur und Fluktuationen (Nobel-Vortrag). Angew. Chem. **90:** 704–715
10 Testa B., Kier L. B. and Carrupt P. A. (1997) A systems approach to molecular structure, intermolecular recognition, and emergence-dissolvence in medicinal research. Med. Res. Rev. **17:** 303–326
11 Koshland D. E. Jr (1976) Role of flexibility in the specificity, control and evolution of enzymes. FEBS Lett. **62:** E47–E52
12 Root-Bernstein R. S. and Dillon, P. F. (1997) Molecular complementarity. I. The complementarity theory of the origin and evolution of life. J. Theor. Biol. **188:** 447–479
13 Campbell D. T. (1974) Downward causation in hierarchically organised systems. In: Studies in the Philosophy of Biology, pp. 179–186, Ayala F.J. and Dobzhansky T. (eds.), Macmillan, London
14 Testa B. and Kier L. B. (2000) Emergence and dissolvence in the self-organization of complex systems. Entropy **2:** 1–25
15 Loewenstein W. R. (2000) The Touchstone of Life: Molecular Information, Cell Communication and the Foundation of Life, Penguin, London
16 Fleissner P. and Hofkirchner W. (1996) Emergent information – towards a unified information theory. Biosystems **38:** 243–248
17 Hruby V. J., Li G., Haskell-Luevano C. and Shenderovich M. (1997) Design of peptides, proteins and peptidomimetics in Chi space. Biopolymers **43:** 219–266
18 Testa B., Raynaud I. and Kier L. B. (1999) What differentiates free amino acids and aminoacyl residues? An exploration of conformational and lipophilicity spaces. Helv. Chim. Acta **82:** 657–665
19 Testa B. and Bojarski A. (2000) Molecules as complex adaptative systems: constrained molecular properties and their biochemical significance. Eur. J. Pharm. Sci. **11 (suppl. 2):** S2– S14

20 Shannon C. (1948) A Mathematical Theory of Communication. Bell Syst. Tech. J. **27:** 379–423, 623–656
21 Fedorov A. A., Magnus K. A. M., Graupe H., Lattman E. E., Pollard T. D. and Almo S. C. (1994) X-ray structures of isoforms of the actin-binding protein profilin that differ in their affinity for phosphatidylinositol phosphates. Proc. Natl. Acad. Sci. USA **91:** 8636–8640
22 Berman H. M., Westbrook J., Feng Z., Gilliland G., Bhat T. N., Weissig H. et al. (2000) The Protein Data Bank. Nucleic Acids Res. **28:** 235–242
23 InsightII User Guide, July 1998, Accelrys Inc., San Diego, Calif.
24 Berendsen H. J. C., Spoel D. van der and Drunen R. van (1995) GROMACS: a message-passing parallel molecular dynamics implementation. Comp. Phys. Comm. **91:** 43–56
25 Lindahl E., Hess B. and Spoel D. van der (2001) GROMACS 3.0: a package for molecular simulation and trajectory analysis. J. Mol. Model. **7:** 306-317
26 Berendsen H. J. C., Postma J. P. M., Gunsteren W. F. van and Hermans J. (1981) Interaction models for water in relation to protein hydration. In: Intermolecular Forces, pp. 331–342, Pullman B. (ed.), Reidel, Dordrecht
27 Berendsen H. J. C., Postma J. P. M., DiNola A. and Haak J. R. (1984) Molecular dynamics with coupling to an external bath. J. Chem. Phys. **81:** 3684–3690
28 Koradi R., Billeter M. and Wüthrich K. (1996) MOLMOL: a program for display and analysis of macromolecular structures. J. Mol. Graphics **14:** 51–55
29 Golomb S. W. (2001) Claude Shannon (1916–2001). Science **292:** 455
30 Stahura F. L., Godden J. W., Xue L. and Bajorath J. (2000) Distinguishing between natural products and synthetic molecules by descriptor Shannon entropy analysis and binary QSAR calculations. J. Chem. Inf. Comput. Sci. **40:** 1245–1252