Journal of Computer-Aided Molecular Design, **15:** 811–817, 2001.
KLUWER/ESCOM

Matching organic libraries with protein-substructures

R. Preissner[a], A. Goede[a], K. Rother[a], F. Osterkamp[b], U. Koert[b] & C. Froemmel[a,*]
[a]*Institute of Biochemistry, Charité, Medical Faculty of the Humboldt-University, Monbijoustr. 2, 10117 Berlin, Germany;* [b]*Institute of Organic Chemistry Humboldt-University, Hessische Str. 3–4, 10115 Berlin, Germany*

Received 29 August 2000; accepted 4 July 2001

Key words: molecular mimicry, peptide mimetic, similarity, superposition, virtual screening

Summary

We present a general approach which allows automatic identification of sub-structures in proteins that resemble given three-dimensional templates. This paper documents its success with non-peptide templates such as β-turn mimetics. We considered well-tested turn-mimetics such as the bicyclic turned dipeptide (BTD), spiro lactam (Spiro) and the 2,5-disubstituded tetrahydrofuran (THF), a new furan-derivative which was recently developed and characterized. The detected geometric similarity between the templates and the protein patches corresponds to r.m.s.-values of 0.3 Å for more than 80% of the constituting atoms, which is typical for active site comparisons of homologous proteins. This fast automatic procedure might be of biomedical value for finding special mimicking leads for particular protein sub-structures as well as for template-assembled synthetic protein (TASP) design.

Introduction

Molecular mimicry as a general principle in biological systems has been proven at various levels of organization. It has been described that DNA [1] or t-RNA [2] are mimicked by peptides and, on the other hand, peptides are functionally imitated by carbohydrate moieties [3–5]. Experimentally validated was the mimicry of a β-turn by a designed α-helix [6], the imitation of the II′ β-turn by D-amino acid containing peptides [7] or the mimicry of viral epitopes by inverse sequence direction [8].

This field has attracted growing attention because the relationship between molecular mimicry and immune-mediated diseases has became increasingly apparent [9]. A major focus of pharmaceutical research has become the structure-based mimicry of original substrates (e.g. immunologically active peptides) by drugs (e.g., peptide-mimetics) [10–12]. Proteomics are leading to a rapid growth of resolved protein structures and rational approaches towards (peptide) ligand design [13, 14]. Because peptides are generally poor drug candidates there is a growing need for chemical peptide substitutes fulfilling geometric criteria such as the secondary structural elements (SSEs) of the mimicked protein [15–17]. The imitation of β-turns draws special attention because of their pre-eminent role in molecular recognition [18]. There are several examples of successful bioactive peptide-mimetic design (for review see [19]), but the proper choice of a template that can replace key side chains is still difficult [20]. A general approach for automatic peptide-substitution by organic compounds should consider their relevant interface (remaining binding properties) rather than rebuilding the entire peptide structure.

It was shown experimentally that a 34-mer peptide mimicking the binding domain of a protein can exhibit similar stability and binding properties [21]. This idea was extended to non-native chain architectures directing peptide blocks to predetermined packing arrangements [22, 23]. The concept of TASP (Template-Assembled Synthetic Proteins) depends on design ideas from modeling systems that are connected to databases of organic peptide substitutes such as CAVEAT or TOPAS [24, 25]. The approach presented here adds direct access to possible fragments of all known protein structures to be screened for simi-

*To whom correspondence should be addressed.
E-mail: cornelius.froemmel@charite.de

larity to potential substitutes. For a number of proteins it was experimentally shown that their folding can be considered as concerted docking of preformed SSEs [26]. This resulted in the idea to use all interior and exterior interfaces of SSEs in proteins as a learning set for molecular recognition, called Dictionary of Interfaces in Proteins (DIP) [27]. Recently, this was applied to detect the building blocks of folding [28] and to examine their evolutionary conservation [29], and also extended to screening for similar interfaces in non-homologous proteins responsible for the binding of the same ligand [30].

Two major applications of DIP are discussed in this paper. First, we address the question of whether interfaces that resemble peptide-mimetics can be detected automatically in a large database of known protein structures to aid the concept of semi-natural proteins (TASP). Secondly, whether the superposition procedure can be applied to test a number of organic lead structures for a suitable imitation of a particular interface. This will be of biomedical value for ligand design in the context of known receptor-substrate-structures [31, 32]. Because the superposition procedure works strictly on a geometric in-variants basis and independently of the chemical nature of the search compound, the screening procedure is unbiased towards amino acid sequence, sequence direction or type of SSE.

Materials and methods

Protein database

A small non-redundant database of 300 proteins was used for this analysis. About 10^4 SSEs make up 10^5 interfaces. Up-to-date Interface files can be downloaded for the 10^4 protein chains of the current 'non-redundancy-list' at the 25% identity level [33]. The database can easily be adapted according to special criteria such as 'all proteins binding flavine-nucleotides' or 'all 4-helix-bundles'. Interface files for further protein structures (as defined in [27]) are available from the authors by email-submission of the structure in PDB-format (confidentiality guaranteed). Database screenings concerning binding similarities, distribution of local packing density or distant (folding) similarities can be performed in collaboration.

Dissection of proteins into exterior and interior patches

The assignment of SSEs takes place according to DSSP [34]. The contact regions (interior patches) are defined using variable distance criterion between the atoms constituting the neighboring SSEs. For details see [27].

Exterior patches are defined via the 'Connolly-surface' [35] of solvent accessible SSEs. The atoms constituting the exterior patch have to be tighter to the Connolly-surface than the cut-off distance. Thus the Connolly-surface represents an ideally packed virtual solvent. All inter-atomic (and to-Connolly) distances smaller than 2.8 Å (largest cut-off) are stored in the Interface files and used for the database generation.

Non-peptide templates

The use of organic templates has a technical, as well as a scientific reason. The technical aspect was to verify that the superposition procedure works independently of the existence of a peptide chain. During pharmaceutical research extensive (experimental) screening efforts are put into finding non-peptide ligands that perform actions similar to their peptide pendants [36]. A theoretical approach that identifies similarities between protein (sub-)structures and organic drug building blocks without manual intervention would be of outstanding scientific value. With this direction in mind we checked a few peptido-mimetics of different chemical type and size (Figure 1). The β-turn (see **1** in Figure 1) is an ubiquitous structural motif involved in many protein recognition processes [37]. The three compounds **2**, **3** and **4** mimic different types of β-turns [38]. The bicyclic turned dipeptide (BTD, [39]) incorporated into the cyclic peptide **2** mimics a β**II**′ turn [40]. The ideal dihedral angles for a β**II**′ turn are $\Phi_2 = +60°$, $\Psi_2 = -120°$, $\Phi_3 = -80°$, $\Psi_3 = 0°$. The NMR structure of the BTD part of **2** in DMSO exhibits a typical a β**II**′ turn conformation ($\Phi_2 = +58°$, $\Psi_2 = -125°$, $\Phi_3 = -67°$, $\Psi_3 = -21°$; [41]). The spiro lactam **3** mimics a β**II** turn [42]. The ideal dihedral angles for a β**II** turn are $\Phi_2 = -60°$, $\Psi_2 = +120°$, $\Phi_3 = +80°$, $\Psi_3 = 0°$. The X-ray structure of **3** displays a typical β**II** turn conformation ($\Phi_2 = -5°$, $\Psi_2 = +129°$, $\Phi_3 = +80°$, $\Psi_3 = -5°$; [42]). The 2,5-disubstituted tetrahydrofuran (THF) dipeptide **4** has an X-ray and NMR-solution structure with the following dihedral angles ($\Phi_2 = -101°$, $\Psi_2 = +52°$, $\omega_2 = +134°$, $\Phi_3 = -113°$, $\Psi_3 = -4°$; [43]). The β-

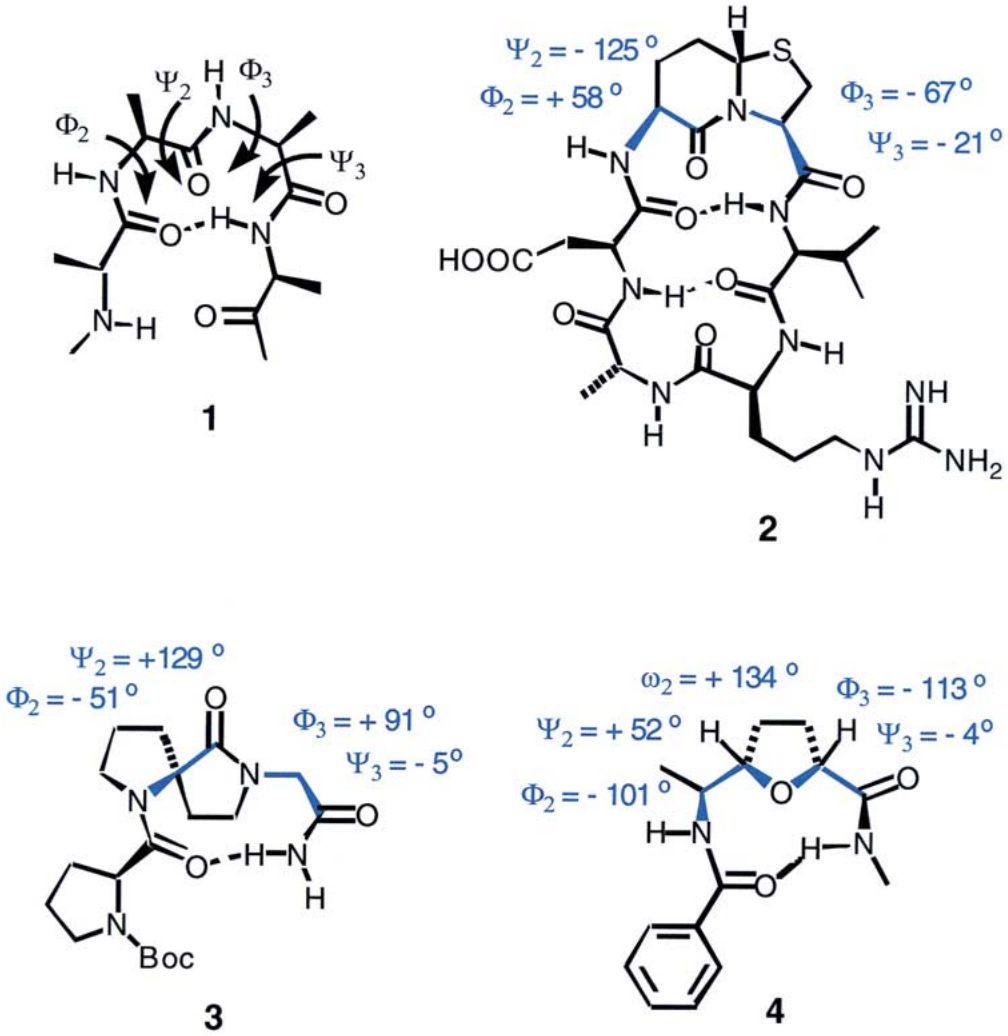

Figure 1. General structure of a β-turn (**1**) and peptido-mimetic templates **2, 3** and **4** used in this analysis.

Table 1. Details concerning the superpositions given in Figure 2

Mimic[a]	Protein[b]	AA[c]		SSE[d]	Comparison in Figure 2	r.m.s. (Å)	Atoms[e]	Chain[f]
2	1thg	389-392		H	a	0.42	12	mc
2	4pep	215-217		C	b	0.37	14	sc
3	2er7	270-272	(E)	H	c	0.55	12	mc
3	1gp1	43-44	(A)	C	d	0.36	12	sc
4	1abm	47-50	(B)	H	e	0.44	12	mc
4	1ton	98-101		C	f	0.39	12	mc

[a]Number of mimic according Figure 1; number of input atoms used: **2** (18), **3** (17), **4** (15).
[b]PDB-code.
[c]Range of superimposed amino acids (peptide chain).
[d]Type of secondary structural element (H: helix, C: coil).
[e]Number of superimposed atoms between template and protein sub-structure.
[f]Majority of superimposed atoms are part of protein backbone (mc) or amino acid side chains (sc).

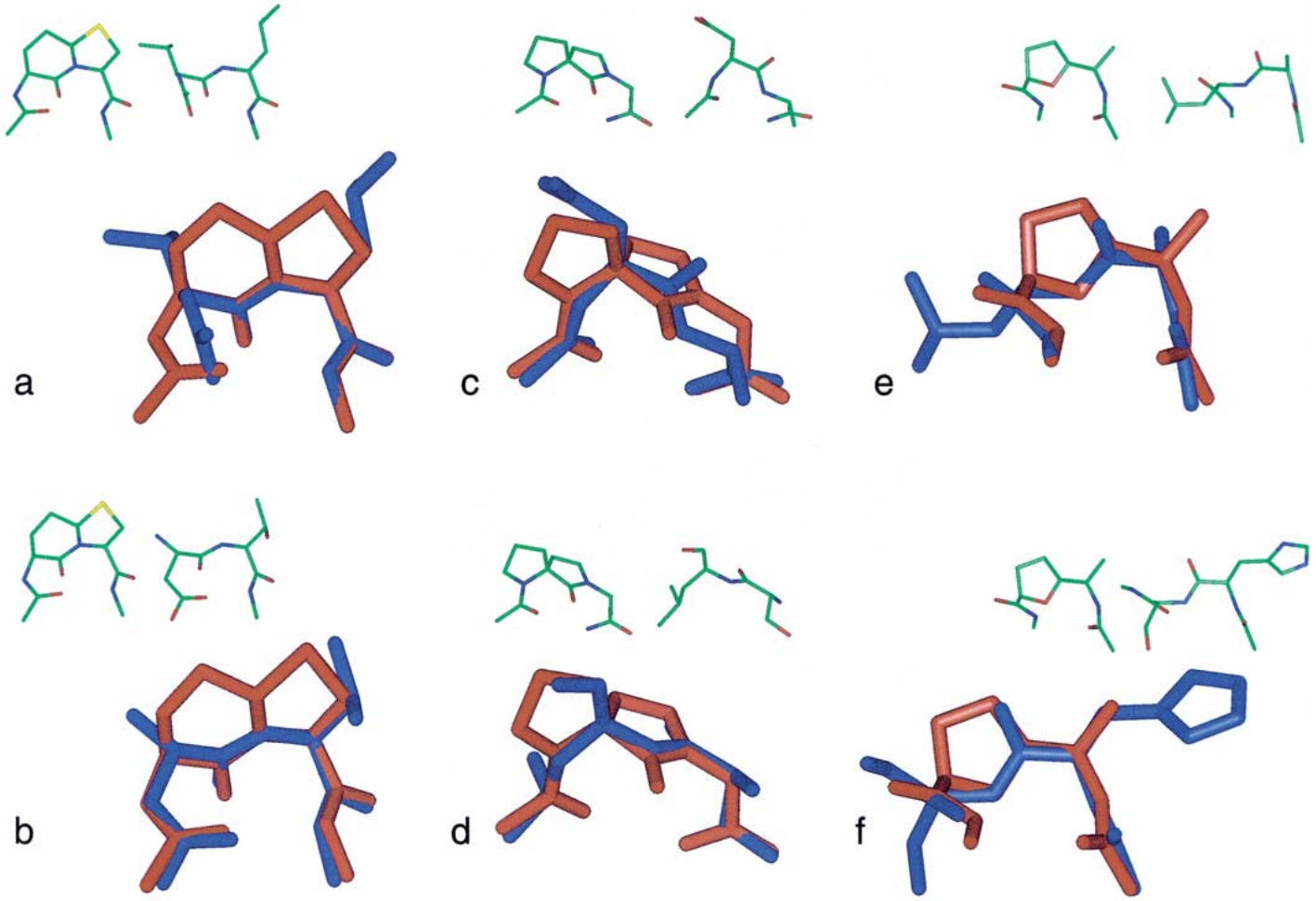

Figure 2. Two of the best alignments for each of the three mimics (67–80% of atoms are aligned at r.m.s.-values of 0.36–0.55 Å). The six superpositions (a–f) are illustrated as follows: the large (lower) part shows the two superimposed structures (mimic colored red, detected protein part in blue). Above both parts (the atoms used for the mimic on the left, the detected amino acids on the right) are given separately at reduced size. They are colored by type of atom (carbon green, oxygen red, nitrogen blue, sulphur yellow). Well-imposed structures with side chain atoms are examples (b) and (d). Because no cis-peptide bond was detected, the ω-angle is not explicitly given for the following examples. (a) Mimic (**1**) aligned with a helical segment of lipase (E.C. 3.1.1.3; PDB-code: 1THG). The indices of the dihedral angles refer to the numbers of the superimposed amino acids: $\Psi_{389} = -47°$; $\Phi_{390} = -61°$; $\Psi_{390} = -44°$; $\Phi_{391} = -67°$. (b) Mimic (**1**) aligned with a loop from Pepsin (E.C. 3.4.23.1; PDB-code: 4PEP). (c) Mimic (**2**) aligned with a helical segment of aspartic proteinase (E.C. 3.4.23.6; PDB-code: 2ER7). The following amino acids contribute to the superposition: Gly270-Asp271-Tyr272. The relevant main chain dihedral angles are $\Psi_{270} = -50°$; $\Phi_{271} = -61°$, $\Psi_{271} = -29°$; $\Phi_{272} = -74°$. (d) Mimic (**2**) matching a loop in glutathione peroxidase (E.C. 1.11.1.9; PDB-code: 1GP1). (e) Mimic (**3**) matching a helical segment (Glu47-Ala48-Leu49) from superoxide dismutase (E.C. 1.15.1.1; PDB-code: 1ABM). The relevant main chain dihedral angles are: $\Psi_{47} = -47°$; $\Phi_{48} = -63°$, $\Psi_{48} = -39°$; $\Phi_{49} = -65°$. (f) Mimic (**3**) matching a coiled segment (Asp98-His99-Ser100) of Tonin (E.C. not assigned; PDB-code: 1TON). The relevant main chain dihedral angles are: $\Psi_{98} = -50°$; $\Phi_{99} = -102°$, $\Psi_{99} = 14°$; $\Phi_{100} = -78°$. For further details on the superposition see Table 1.

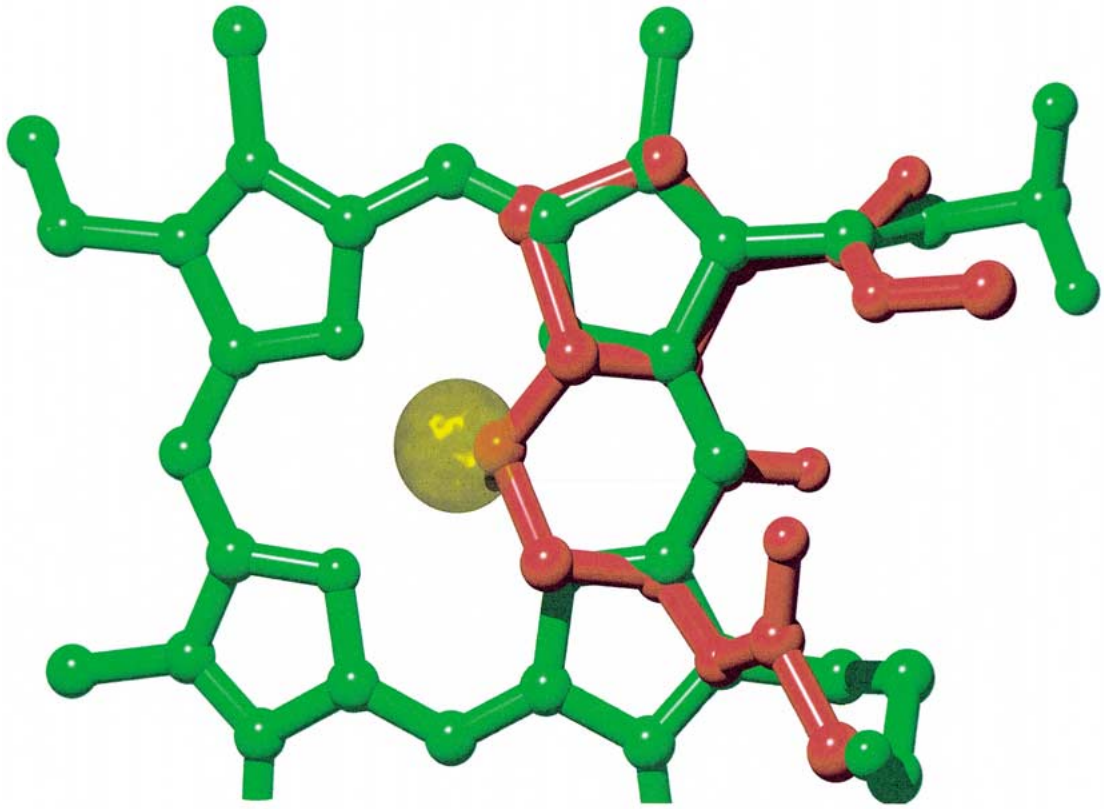

Figure 3. Accidental match between a (part of a) hem hetero-cycle (green) in hemoglobin (PDB-code: 2HBG) and the search template **2** at r.m.s.-deviation of 0.54 Å for 13 (of 18) atoms. The complex with the Fe-atom is shown as transparent yellow sphere, while mimic 1 is drawn by red sticks.

turn type of **4** is difficult to fit into a known turn-type category based on its dihedral angles.

Similarity screening

The similarity screening parameters were adjusted such that more than two thirds of all atoms match at r.m.s.-values below 1.0 Å. These parameters were used for the screening: i.e., for a mimic of 21 atoms (mimic 1) all patches between 14 and 28 atoms have to taken into account (4×10^5 patches). The extent of the search template (7.8 Å, 6.7 Å, 3.4 Å) $\pm$ 1 Å defines a possible search space. The use of a database patch-list presorted according to the lengths, widths and depths of patches further reduces the number of considered patches in the range [(6.8–8.8 Å), (5.7–7.7 Å), (2.4–4.4 Å)] to 3×10^3. The resulting 10^4 comparisons can be performed on an IBM PC in a few minutes. For details concerning the superposition algorithm see [27]. The superposition procedure for two patches is included in a viewer that can be downloaded with the Interface files.

Results and discussion

Screening with the peptido-mimetics described in Materials and Methods revealed about one hundred patches with significant similarity. We analyzed the patches with respect to the following aspects: root mean square deviation (r.m.s.), number of superimposed atoms; type of original SSE, type of neighboring SSE; quality of backbone or side chain match.

These data are compiled in Table 1 for the three mimics. For each search compound two superpositions with protein patches are visualized in Figure 2. Although the search compounds are 'turn-mimetics', solvent directed patches occurred, but did not dominate the screening results. The neighboring patch may be of interest for the binding properties of the mimic found. If the binding properties are of particular importance, the superposition of the side chain atoms should be optimized (Figure 2b, d). Curiously, we found a reasonable superposition between part of an organic hetero-cycle (hem) and one of the mimics

(see Figure 3). This finding was a side-effect, demonstrating the potential of the method to detect new ligands.

The idea to incorporate non-native chain architectures into peptides was realized for an analogue of vasopressin by incorporation of a turn-mimetic [44] and for mimic 3 of this analysis as part of a zinc-finger [45]. This concept of TASP will be directly aided through design ideas by the automatic screening procedure presented here.

In summary, it was demonstrated that the Dictionary of Interfaces in Proteins (DIP, [27]) works for non-peptide compounds. Using the included superposition algorithm it will be possible to screen databases of natural products, pharmacological leads or arbitrary low-molecular weight compounds (e.g., Cambridge Structural Database; [46]) against all known protein structures and vice versa.

Prospects

An optimization criterion during molecular evolution may not only be binding of specific substrates or co-factors, but the non-binding to numerous substances, including other proteins. To understand how this works it would be helpful to classify all exterior patches according to geometric similarity, as was initiated for helix-helix interfaces [28]. Unbiased screening will then be possible, detecting similarities to protein surfaces that might be responsible for binding or immuno-relevant (cross-)recognition.

References

1. Putnam, C.D., Shroyer, M.J., Lundquist, A.J., Mol, C.D., Arvai, A.S., Mosbaugh, D.W. and Trainer, J.A., J. Mol. Biol., 287 (1999) 331.
2. Ito, K., Ebihara, K., Uno, M. and Nakamura, Y., Proc. Natl. Acad. Sci. USA, 93 (1996) 5443.
3. Kaur, K.J., Khurana, S. and Salunke, D.M., J. Biol. Chem., 272 (1997) 5539.
4. Verma, S. and Eckstein, F., Annu. Rev. Biochem., 67 (1998) 99.
5. Sears, P. and Wong, C.H., Angew. Chem. Int. Ed., 38 (1999) 2300.
6. Mer, G., Kellenberger, E. and Lefevre, J.F., J. Mol. Biol., 281 (1998) 235.
7. Weisshoff, H., Prasang, C., Henklein, P., Froemmel, C., Zschunke, A. and Muegge, C., Eur. J. Biochem., 259 (1999) 776.
8. Benkirane, N., Guichard, G., Briand, J.P., Muller, S., Brown, F. and van-Regenmortel, M.H., Dev. Biol. Stand., 87 (1996) 283.
9. Oldstone, M.B., FASEB-J., 12 (1998) 1255.
10. Murali, R. and Greene, M.I., Immunol. Res., 17 (1998) 163.
11. Reineke, U. and Schneider-Mergener, J., Angew. Chem. Int. Ed. Engl., 37 (1998) 3241.
12. Lehmann, P.V., Trends Pharmacol. Sci., 21 (2000) 79.
13. Patel, S., Scott, I.P., Bhakoo, M. and Elliott, P., J. Comput. Aid. Mol. Des., 12 (1998) 543.
14. Schneider, G., Neidhart, W., Giller, T. and Schmid, G., Angew. Chem. Int. Ed., 38 (1999) 2894.
15. Mueller, G. and Giera, H., J. Comput. Aid. Mol. Des., 12 (1998) 1.
16. Smith, M.D. and Fleet, G.W., J. Pept. Sci., 5 (1999) 425.
17. Stigers, K.D., Soth, M.J. and Nowick, J.S., Curr. Opin. Chem. Biol., 3 (1999) 714.
18. Garland, S.L. and Dean, P.M., J. Comput. Aid. Mol. Des., 13 (1999) 469.
19. Ripka, A.S. and Rich, D.H., Curr. Opinion in Chem. Biol., 2 (1998) 441.
20. Hruby, V.J. and Balse, P.M., Curr. Med. Chem., 7 (2000) 945.
21. Starovasnik, M.A., Braisted, A.C. and Wells, J.A., Proc. Natl. Acad. Sci. USA, 94 (1997) 10080.
22. Mutter, M. and Tuchscherer, G., Cell. Mol. Life Sci., 53 (1997) 851.
23. Tuchscherer, G., Grell, D., Mathieu, M. and Mutter, M., J. Pept. Res., 54 (1997) 185.
24. Lauri, G. and Bartlett, P.A., J. Comput. Aid. Mol. Des., 8 (1994) 51.
25. Schneider, G., Lee, M.-L., Stahl, M., Schneider, P., J. Comput. Aid. Mol. Des., 14 (2000) 487.
26. Peng, Z.Y., Wu, L.C., Schulman, B.A. and Kim, P.S., Phil. Trans. Roy. Soc. Ser., **B348** (1996) 43.
27. Preissner, R., Goede, A. and Frömmel, C., J. Mol. Biol., 280 (1998) 535.
28. Preissner, R., Goede, A. and Frömmel, C., Protein Eng., 12 (1999) 825.
29. Gille, C., Goede, A., Preissner, R., Rother, K. and Frömmel, C., J. Mol. Biol., 299 (2000) 1147.
30. Preissner, R., Goede, A. and Frömmel, C., Bioinformatics, 15 (1999) 832.
31. Preissner, R., Goede, A. and Frömmel, C., European patent 97 928 126.8, 1997.
32. Reineke, U., Preissner, R., Goede, A., Germeroth, L., Schneider-Mergener, J. and Frömmel, C., European Peptide Symposium, Montpellier, France, Sept. 10-15, 2000.
33. Hobohm, U. and Sander, C., Protein Sci., 3 (1994) 522.
34. Kabsch, W. and Sander, C., Biopolymers, 22 (1983) 2577.
35. Connolly, M.L., J. Appl. Crytallogr., 16 (1983) 548.
36. Boehm, H-J. and Klebe, J., Angew. Chem. Int. Ed. Engl., 35 (1996) 2588.
37. Rose, G.D., Gierasch, L.M. and Smith, J.A., Adv. Protein Chem., 37 (1985) 1.
38. Ball, J. B., Hughes, R. A., Alewood, P. F. and Andrews, P. R., Tetrahedron, 49 (1993) 3467.
39. Nagai, U., Sato, K., Nakamura, R. and Katao, R., Tetrahedron, 49 (1993) 3577.
40. Doyle, P. M., Harris, J. C. Moody, C. M., Sadler, P. J., Sims, M., Thornton, J. M., Uppenbrink, J. and Viles, J. H., Int. J. Peptide Protein Res., 47 (1996) 427.
41. Haubner, R. Schmitt, W., Hoelzemann, G., Goodman, S. L., Jonczyk, A. and Kessler, H., J. Am. Chem. Soc., 118 (1996) 7881.
42. Genin, M. J., Ojala, W. H., Gleason, W. B. and Johnson, R. L., J. Org. Chem. 58 (1993) 2334.
43. Schrey, A., Osterkamp, F., Straudi, A., Rickert, C., Wagner, H., Koert, U., Herrschaft, B. and Harms, K., Eur. J. Org. Chem., 47 (1999) 2977.

44. Brickmann, K., Yuan, Z., Sethson, I., Somfai, P. and Kihlberg, J., Chem. Eur. J., 5 (1999) 2241.
45. Viles, J.H., Patel, S.U., Mitchell, J. B. O., Moody, C. M., Justice, D. E., Uppenbrink, J., Doyle, P.M., Harris, C. J., Sadler, P.J. and Thornton, J. M., J. Mol. Biol., 279 (1998) 973.
46. Allen, F.H. and Kennard, O., Perspect.Comput., 3 (1983) 28.